

COMISSÃO EUROPEIA PARA A
EFICIÊNCIA DA JUSTIÇA (CEPEJ)

Carta Ética Europeia sobre o uso da
Inteligência Artificial em sistemas
judiciais e seu ambiente



Charte éthique européenne d'utilisation de
intelligence artificielle dans les systèmes
judiciaires et leur environnement



Foz do Iguaçu - 2019

Painel VI: Inteligência Artificial

Presidente de Mesa:
Conselheiro Márcio Schiefler Fontes Conselho
Nacional de Justiça - CNJ

Debatedores:
Conselheira Maria Tereza Uille
Conselho Nacional de Justiça – CNJ

Des°. Marcos Alaor Grangeia
coordenador da Escola Nacional da Magistratura - ENM

Des°. Samuel Meira Brasil Junior
coordenador da Escola Nacional da Magistratura – ENM

Charte éthique européenne d'utilisation de intelligence artificielle dans les systèmes judiciaires et leur environnement

Tradução livre de:

Teresa Germana Lopes de Azevedo

Magistrada do Poder Judiciário do Ceará



Carta Ética Europeia sobre o uso da
Inteligência Artificial em sistemas
judiciais e seu ambiente

CONTEÚDO

INTRODUÇÃO

OS CINCO PRINCÍPIOS DA CARTA ÉTICA SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL EM SISTEMAS JUDICIAIS E SEU AMBIENTE

1. Princípio do respeito pelos direitos fundamentais: assegurar que a elaboração e a implementação de ferramentas e serviços de Inteligência Artificial sejam compatíveis com os direitos fundamentais
2. Princípio da não discriminação: Prevenir o desenvolvimento ou intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos
3. Princípio da qualidade e segurança: No processamento de decisões e dados judiciais, utilizar fontes certificadas e dados intangíveis com modelos elaborados de forma multidisciplinar, num ambiente tecnológico seguro
4. Princípio da transparência, imparcialidade e justiça: tornar os métodos de processamento de dados acessíveis e compreensíveis, autorizar auditorias externas
5. Princípio “sob controle do usuário”: impedir uma abordagem prescritiva e garantir que os usuários sejam informados e controlem suas escolhas

ANEXO I - ESTUDO APROFUNDADO SOBRE O USO DA IA EM SISTEMAS JUDICIAIS, INCLUINDO APLICAÇÕES, GARANTINDO O PROCESSAMENTO DE DECISÕES E DADOS JUDICIAIS

1. Estado da utilização de algoritmos de inteligência artificial nos sistemas judiciais dos Estados membros do Conselho da Europa
2. Situação das políticas públicas de decisões de dados abertos nos sistemas judiciais dos Estados membros do Conselho da Europa
3. Características operacionais da inteligência artificial (aprendizado de máquina) aplicada à decisão judicial
4. A inteligência artificial pode modelar o raciocínio jurídico?
5. As IA's podem explicar o comportamento dos juízes a posteriori?
6. Como a IA deve ser aplicada na justiça civil, comercial e administrativa?

7. Questões específicas da justiça criminal: prevenção de infrações, risco de reincidência e avaliação de periculosidade
8. Questões específicas relacionadas com a proteção de dados pessoais
9. O potencial e as limitações das ferramentas de justiça preditiva
10. A necessidade de um debate público aprofundado sobre estas ferramentas antes da implementação de políticas públicas para o seu desenvolvimento. A necessidade urgente da ciberética para o desenvolvimento de algoritmos de inteligência artificial, respeitando os direitos fundamentais

ANEXO II – QUAIS OS USOS DA IA NOS SISTEMAS JUDICIAIS EUROPEUS?

Utiliza-se para encorajar

Possíveis usos, exigindo precauções metodológicas

Usos a serem considerados após estudos científicos adicionais

Usos a considerar com reservas mais extremas

ANEXO III: GLOSSÁRIO

ANEXO IV: CHECKLIST PARA INTEGRAR OS PRINCÍPIOS DA CARTA NOS SEUS TRATAMENTOS

CHECHLIST PARA AVALIAR SEUS MÉTODOS DE PROCESSAMENTO

Introdução

Reconhecendo a crescente importância da inteligência artificial¹ (IA) em nossas sociedades modernas, e os benefícios esperados quando totalmente utilizada para eficiência e qualidade da justiça, o CEPEJ adota formalmente os cinco princípios fundamentais intitulados “Carta Ética Europeia sobre o uso da IA nos sistemas judiciais e seu ambiente”.

Esta Carta é dirigida aos atores públicos e privados responsáveis pelo projeto e implantação de ferramentas de inteligência artificial e serviços que envolvam o processamento de decisões e dados judiciais (aprendizado de máquina ou qualquer outro método derivado da ciência de dados).

Refere-se igualmente aos agentes públicos responsáveis pelo processo legislativo ou regulamentar, desenvolvimento, auditoria ou uso de tais ferramentas e serviços.

A utilização de tais ferramentas e serviços nos sistemas judiciais destina-se a melhorar a eficiência e qualidade da justiça e deve ser encorajada. No entanto, isso deve ser feito com responsabilidade, respeitando os direitos fundamentais dos indivíduos, tal como previsto na Convenção Europeia dos Direitos Humanos e na Convenção para a Proteção de Dados Pessoais, e em conformidade com outros princípios estabelecidos abaixo, que devem orientar a definição de políticas públicas de justiça nesse campo.

O processamento de decisões judiciais por inteligência artificial, de acordo com seus desenvolvedores, provavelmente, em questões civis, comerciais e administrativas, ajudará a melhorar a previsibilidade da aplicação da lei e a coerência das decisões judiciais, sob reserva do cumprimento dos princípios abaixo enunciados. Em matéria penal, seu uso deve ser considerado com as maiores reservas, a fim de evitar a discriminação com base em dados sensíveis, em conformidade com as garantias de um julgamento justo.

Devem ser concebidos com o objetivo de auxiliar na prestação de consultoria jurídica, auxiliando na elaboração ou no processo de tomada de decisões, ou aconselhando o usuário. É essencial que o processamento seja realizado com transparência, imparcialidade e equidade, certificado por uma perícia, externa ao operador, e independente.

¹ Para a definição de inteligência artificial, consulte o Glossário em anexo.

Aplicação da Carta

Os princípios da Carta devem ser sujeitos a uma aplicação, monitorização e avaliação regular pelos atores públicos e privados, tendo em vista a melhoria das práticas.

A este respeito, é desejável que uma revisão regular da implementação dos princípios da Carta seja feita por esses atores, explicando, se necessário, as razões da não implementação ou da execução parcial, acompanhada de um plano de ação para introduzir as medidas necessárias.

As autoridades independentes mencionadas na Carta poderiam ser responsáveis por avaliar periodicamente o nível de adesão aos princípios desta Carta por todos os atores e propor melhorias para se adaptar às novas tecnologias e usos dessas tecnologias.

OS CINCO PRINCÍPIOS DA CARTA ÉTICA SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL EM SISTEMAS JUDICIAIS E SEU AMBIENTE

1. Princípio do respeito pelos direitos fundamentais: assegurar que a elaboração e a implementação de ferramentas e serviços de Inteligência Artificial sejam compatíveis com os direitos fundamentais
2. Princípio da não discriminação: Prevenir o desenvolvimento ou intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos
3. Princípio da qualidade e segurança: No processamento de decisões e dados judiciais, utilizar fontes certificadas e dados intangíveis com modelos elaborados de forma multidisciplinar, num ambiente tecnológico seguro
4. Princípio da transparência, imparcialidade e justiça: tornar os métodos de processamento de dados acessíveis e compreensíveis, autorizar auditorias externas
5. Princípio “sob controle do usuário”: impedir uma abordagem prescritiva e garantir que os usuários sejam informados e controlem suas escolhas

1. Princípio do respeito pelos direitos fundamentais: assegurar que a elaboração e a implementação de ferramentas e serviços de Inteligência Artificial sejam compatíveis com os direitos fundamentais

- O processamento de decisões e dados judiciais deve ter objetivos claros, em plena conformidade com os direitos fundamentais garantidos pela Convenção Europeia dos Direitos Humanos (CEDH) e a Convenção para a Proteção de Dados Pessoais (Convenção para a Proteção de Indivíduos com relação ao Processamento Automático de Dados Pessoais, ETS N° 108, emendada pelo protocolo de alteração n ° 223).
- Quando ferramentas de inteligência artificial são usadas para resolver uma disputa, ou como uma ferramenta para auxiliar na tomada de decisões judiciais ou dar orientação ao público, é essencial garantir que não prejudiquem as garantias do direito de acesso ao juiz e o direito a um julgamento justo (igualdade de armas e respeito ao contraditório).
- Elas também e devem ser usadas com o devido respeito aos princípios do Estado de direito e da independência dos juízes em seu processo decisório.
- Portanto, deve-se dar preferência as abordagens chamadas “ética da concepção²” ou “direitos humanos desde a concepção”. Elas integram, a partir da fase de projeto e aprendizagem, regras que proíbem a violação direta ou indiretamente de valores fundamentais protegidos pelas Convenções.

² A escolha ética é feita pelos projetistas do programa e, portanto, não é deixada para o usuário.

2. Princípio da não discriminação: Prevenir o desenvolvimento ou intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos

- Dada a capacidade destes métodos de processamento de revelar a discriminação existente, através do agrupamento ou classificação de dados relativos a indivíduos ou grupos de indivíduos, as partes interessadas, públicas e privadas, devem garantir que os métodos não irão reproduzir ou agravar estas discriminações e que não irão conduzir à análises ou usos deterministas.
- Cuidados especiais devem ser tomados tanto no desenvolvimento quanto na fase de implantação, especialmente quando o processamento é direta ou indiretamente baseado em dados “sensíveis”. Isso pode incluir origem racial ou étnica, condições socioeconômicas, opiniões políticas, crenças religiosas ou filosóficas, adesão sindical, dados genéticos, dados biométricos, dados relacionados com a saúde ou dados relativos à vida sexual ou orientação sexual. Quando tal discriminação tem sido identificada, devem ser consideradas as medidas corretivas para limitar ou, se possível, neutralizar esses riscos, bem como a conscientização das partes interessadas deve ser considerada.
- No entanto, o uso do aprendizado de máquina e a realização de análises científicas multidisciplinares para combater essa discriminação devem ser encorajados.

3. Princípio da qualidade e segurança: No processamento de decisões e dados judiciais, utilizar fontes certificadas e dados intangíveis com modelos elaborados de forma multidisciplinar, num ambiente tecnológico seguro

- Designers de modelo da aprendizagem devem ser capazes de coletar de uma maneira ampla a perícia dos profissionais da justiça em questão (juízes, promotores, advogados, etc.) e pesquisadores / professores de direito e ciências sociais (economistas, sociólogos, filósofos, por exemplo).
- A constituição de equipes mistas de projetos em ciclos de designe curto para produzir modelos funcionais é um dos modos de organização que permite obter a melhor parte desta multidisciplinaridade.
- As medidas éticas existentes devem poder ser constantemente compartilhadas por essas equipes de projeto e aprimoradas com base no feedback.
- Dados baseados em decisões judiciais inseridas em um software que implementa um algoritmo de aprendizado de máquina deve vir de fontes certificadas e não devem ser modificados até que eles tenham sido utilizados de forma eficaz pelo mecanismo de aprendizagem. Todo o processo deve, portanto, ser rastreável para garantir que nenhuma modificação ocorreu para alterar o conteúdo ou significado da decisão processada.
- Os modelos e algoritmos criados também devem poder ser armazenados e executados em ambientes seguros, de modo a garantir a integridade do sistema e sua intangibilidade.

4. Princípio da transparência, imparcialidade e justiça: tornar os métodos de processamento de dados acessíveis e compreensíveis, autorizar auditorias externas

- Um equilíbrio deve ser encontrado³ entre a propriedade intelectual de determinados métodos de processamento e a necessidade de transparência (acesso ao projeto de design), imparcialidade (ausência de preconceito⁴), justiça e integridade intelectual (priorizando os interesses da justiça), devido a efeitos legais ou impactos que esses métodos podem ter na vida das pessoas. Deve-se notar que essas medidas se aplicam a todo o projeto e à cadeia operacional, desde o processo de seleção e a qualidade e organização dos dados, que influenciam diretamente a fase de aprendizagem.
- A primeira opção é a transparência técnica completa (por exemplo, fonte aberta de código e documentação), às vezes restrita pela proteção de segredos comerciais. O sistema também pode ser explicado de forma clara e linguagem familiar (para descrever como os resultados são produzidos) comunicando, por exemplo, a natureza dos serviços oferecidos, as ferramentas desenvolvidas, o desempenho e os riscos de erro. Autoridades independentes ou especialistas poderiam ser encarregados de certificar e auditar métodos de processamento ou fornecer aconselhamento. As autoridades públicas podem conceder certificação.

³ De interesse nesta conexão é a sugestão feita na página 38 do Conselho da Europa. A proposta apresentada no estudo MSI-NET intitulado "Algoritmos e Direitos Humanos": "A solução de divulgação pública de algoritmos inteiros ou de seu código-fonte é improvável neste contexto, empresas privadas consideram seu algoritmo como software proprietário estratégico e, portanto, devem protegê-los adequadamente. Contudo, é possível exigir a publicação de informações parciais, mas importantes, como, por exemplo, os objetivos de otimização de algoritmos, os conjuntos de dados de aprendizagem, os valores médios e desvios-padrão dos resultados obtidos e a quantidade e tipo de dados processados pelo o algoritmo". Até mesmo as sugestões que aparecem na página 117 do relatório intitulado "IA para a humanidade", redigido pelo Sr. Cédric Villani, um membro da Assembleia Nacional francesa como parte de uma missão designada para ele pelo Primeiro Ministro da República Francesa: "Os auditores podem ficar satisfeitos em verificar a justiça e equidade de um programa (fazendo o que se supõe que seja feito), através do envio de vários dados de entrada falsos, por exemplo, ou criando uma grande quantidade de perfis de usuário de acordo com diretrizes precisas". Além disso, há também as observações no relatório da Câmara dos Lordes, "IA no Reino Unido: pronto, disposto e capaz?", parágrafos 92, 96-99.

⁴ Neste contexto, é interessante notar as soluções geralmente consideradas para garantir a neutralidade dos algoritmos no relatório acima mencionado pela Câmara dos Lordes (parágrafos 114, 115, 116, 119, 120): conjuntos de dados mais diversificados, mais diversidade e multidisciplinaridade de abordagens, mais auditoria de aspectos como processamento de dados e a maneira como a máquina é construída.

5. Princípio “sob controle do usuário”: impedir uma abordagem prescritiva e garantir que os usuários sejam informados e controlem suas escolhas

- A autonomia do usuário deve ser reforçada e não restrita pelo uso de ferramentas e serviços de inteligência artificial.
- Os profissionais do sistema judiciário devem, a qualquer momento, poder rever as decisões judiciais e os dados utilizados para produzir um resultado e continuar a ter a oportunidade de partir dela, de acordo com as especificidades do caso concreto.
- O usuário (litigante) deve ser informado em linguagem clara e compreensível se as soluções oferecidas pelas ferramentas de inteligência artificial são vinculativas ou não as diferentes opções disponíveis, bem como do seu direito à assistência jurídica e de acessar um tribunal. O usuário (litigante) também deve ser claramente informado de qualquer processamento de um caso por inteligência artificial antes ou durante um processo judicial e tem o direito de se opor, para que seu caso possa ser ouvido diretamente por um tribunal, na acepção do artigo 6º da CEDH.
- De um modo geral, quando qualquer sistema de informação baseado em inteligência artificial é implementado deve haver programas de informática para usuários e debates envolvendo profissionais do sistema de justiça.

ANEXO I - ESTUDO APROFUNDADO SOBRE O USO DA IA EM SISTEMAS JUDICIAIS, INCLUINDO APLICAÇÕES, GARANTINDO O PROCESSAMENTO DE DECISÕES E DADOS JUDICIAIS

Preparado pelo Sr. Xavier Ronsin, Primeiro Presidente do Tribunal de Recurso do Rennes, especialista científico (França), e Vasileios Lampos, pesquisador Sênior do departamento de ciências da computação da University College London (UCL), especialista científico (Reino Unido),

e com a contribuição de Agnès Maîtrepierre, juíza, membro do Comitê Consultivo da Convenção para a Proteção das Pessoas Singulares no que diz respeito ao Processamento Automático de Dados Pessoais do Conselho da Europa (França)

Os seguintes especialistas também contribuíram para o estudo:

Francesco Contini, pesquisador sênior do Instituto de Pesquisa em Sistemas Judiciais - Centro Nacional de Pesquisa (IRSIG-CNR), Bolonha (Itália)

Francesco De Santis, professor de Procedimentos para a Proteção dos Direitos Humanos, Universidade de Nápoles (Itália)

Jean Lassègue, filósofo e epistemólogo, pesquisador do Centro Nacional de Pesquisa Científica (CNRS) e pesquisador associado ao Instituto de Estudos Superiores em Justiça (IHEJ) (França)

Dory Reiling, Juiz Honorário Sênior, Especialista Independente em Tecnologias da Informação e em reformas judiciais (Holanda)

Aleš Završnik, Pesquisador Chefe do Instituto de Criminologia, Professor associado da Faculdade de Direito da Universidade de Ljubljana (Eslovénia) e Pesquisador EURIAS 2017-18 no Collegium Helveticum em Zurique (Suíça)

INTRODUÇÃO

1. A onda de transformação digital de nossas sociedades ainda é desigual entre os sistemas judiciais dos Estados membros do Conselho da Europa. Muitos países europeus já parecem ter desenvolvido um pensamento extremamente avançado com aplicações em concreto (tanto tecnologicamente, quanto em termos de apoio jurídico), enquanto para outros esta ainda é apenas uma questão emergente e o foco é exclusivamente no gerenciamento eficiente da Tecnologia de Informação.

2. Entre as tecnologias em ação nessa grande transformação digital, a inteligência artificial (IA) parece ser a mais espetacular. Nos Estados Unidos, “advogados robôs” já estão trabalhando e parecem conversar em linguagem natural com humanos. Startups de tecnologia legal especializadas no design de novos serviços jurídicos oferecem novas aplicações para profissionais, principalmente advogados, serviços jurídicos e seguradoras, permitindo acesso a informação judicial e a jurisprudência. Essas empresas privadas visam prever as decisões dos juízes com ferramentas de “justiça preditiva”, embora essa não seja a melhor nomenclatura⁵.

3. Um exame inicial desse fenômeno, entretanto, leva à distinção entre o discurso comercial e a realidade do uso e implantação dessas tecnologias. Os juízes dos Estados membros do Conselho da Europa não parecem estar fazendo uso diário de software preditivo. Experiências locais⁶ e trabalho acadêmico⁷ foram realizadas para explorar o potencial dessas aplicações, mas elas ainda não foram aplicadas em larga escala. A iniciativa para o desenvolvimento destas ferramentas vem em grande parte do setor privado, cuja clientela até agora é composta principalmente por companhias de seguros, advogados e serviços jurídicos comprometidos com a redução do “risco judicial”, isto é, a imprevisibilidade da decisão judicial. No entanto, começam a ser cada vez mais solicitados por um setor privado que deseja ver essas ferramentas - que às vezes são versões “beta”, ou seja, elas evoluirão ao longo do tempo - integradas

⁵ Ver, em especial, o quadro no capítulo 9.

⁶ Por exemplo, os Tribunais de Apelação de Douai e Rennes, na França, realizaram um julgamento em 2017 com um programa de software rotulado como “preditivo” por um painel de juizes.

⁷ Trabalho com uma amostra de 584 decisões do Tribunal Europeu dos Direitos do Homem: Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiu-Pietro, Vasileios Lampos, “Prevedendo decisões judiciais da Corte Europeia de Direitos Humanos: Uma Perspectiva do Processamento de Linguagem Natural”, publicado em 24 de outubro de 2016, [Online], <https://peerj.com/articles/cs-93/>

às políticas públicas.

4. Na linha do processo de pensamento iniciado em suas “Diretrizes sobre como conduzir a mudança para a Cyberjustice⁸”, o CEPEJ propõe oferecer aos profissionais da justiça a chave para uma melhor compreensão do fenómeno “Justiça preditiva”.

5. A primeira tarefa será esclarecer as questões relativas à natureza intrínseca desses sistemas de processamento de dados de jurisprudência em massa, bem como suas limitações técnicas e teóricas. Estes aspectos não têm sido frequentemente mencionados no debate sobre este assunto na esfera judicial, mas eles são bem conhecidos e discutidos por especialistas nessas tecnologias (matemáticos, estatísticos e cientistas da computação) e merecem atenção.

6. Em segundo lugar, este documento analisará os benefícios e riscos dessas Ferramentas. Enquanto seus defensores destacam seus pontos fortes em termos de transparência, previsibilidade e padronização da jurisprudência, seus críticos apontam para as limitações e o viés de raciocínio do software atualmente no mercado. Os riscos inerentes a essas tecnologias podem até transcender o ato de julgar e afetar elementos essenciais ao funcionamento do Estado de direito e dos sistemas judiciais, aos quais o Conselho da Europa está particularmente ligado.

7. Estes incluem princípios como o Estado de direito. O efeito destas ferramentas podem não ser apenas para fornecer incentivos, mas para ser quase prescritiva, criando uma nova forma de normatividade, que poderia complementar a lei regulando a discricção soberana do juiz e, potencialmente, liderando, a longo prazo, para uma padronização das decisões judiciais baseadas não mais no raciocínio judicial individualizado, mas em um cálculo estatístico puro vinculado à média anteriormente atribuída por outros tribunais.

8. Há também a necessidade de considerar se estas soluções são compatíveis com os direitos individuais consagrados na Convenção Europeia dos Direitos Humanos (CEDH). Estes incluem o direito a um julgamento justo (particularmente o direito a um juiz natural estabelecido por lei, o direito a um tribunal independente e imparcial e a igualdade de armas em processos judiciais) e o direito ao respeito pela vida privada e familiar, particularmente o

⁸ Ver, em particular, o parágrafo 51 do documento CEPEJ (2016) 13, Diretrizes sobre como conduzir para mudar para Cyberjustice.

respeito pela proteção dos dados comunicados em dados abertos.

9. Embora sensível a essas questões, o documento observa o imenso potencial da IA como um apoio ao trabalho dos profissionais da justiça. Não há dúvida que algumas aplicações da IA que ainda estão em desenvolvimento ou em teste, como aquelas projetadas para melhorar a pesquisa jurídica, possam ser muito úteis e tornar o trabalho judicial mais rápido e mais eficiente. O documento destaca esses exemplos positivos e defende o uso da IA por profissionais da área jurídica de acordo com suas necessidades, desde que seja dada especial atenção aos direitos individuais garantidos pelas normas da CEDH e do Conselho da Europa, particularmente em matéria penal. Longe de ser um instrumento simples para melhorar a eficiência dos sistemas judiciais, a IA deve reforçar as garantias do Estado de direito, juntamente com a qualidade da justiça pública.

10. Por fim, o documento dá algumas ideias para uma “governança” deste fenômeno, na forma de uma carta ética, enfatizando a necessidade de uma abordagem para a integração dessas ferramentas nas políticas públicas. É essencial que qualquer debate público envolva todas as partes interessadas, sejam profissionais de direito, empresas jurídicas de tecnologia ou cientistas, para permitir-lhes compartilhar sua análise sobre o escopo e o possível impacto da introdução de aplicativos de IA no sistemas judiciais e, também, conceber o quadro ético em que eles devem operar. Posteriormente, este debate poderia ir além de um puro “negócio”, envolvendo os próprios cidadãos e, portanto, contribuindo de certa forma para a alfabetização digital, como tem sido feito, por exemplo, no Canadá⁹.

⁹ Declaração de Montreal, disponível em <https://www.declarationmontreal-iaresponsable.com/> / demarche, acessado em 16 de julho de 2018.

1. Estado da utilização de algoritmos de inteligência artificial nos sistemas judiciais dos Estados membros do Conselho da Europa

Em 2018, o uso de algoritmos de inteligência artificial nos sistemas judiciais europeus permanece principalmente por iniciativa do setor privado, voltado para seguradoras, departamentos jurídicos, advogados e indivíduos.

11. O uso da IA no campo judicial parece ser bastante popular nos Estados Unidos, que investiu bastante nessas ferramentas, tanto em matéria civil como penal¹⁰.

12. A Identificação de iniciativas de algoritmos de IA nos sistemas judiciais dos Estados membros do Conselho da Europa é uma tarefa mais difícil, uma vez que a maioria das iniciativas vem do setor privado e muitas vezes não são integradas às políticas públicas.

13. A questão da utilização da IA nos sistemas judiciais foi abordada em uma pesquisa online específica, lançada em abril de 2018 para representantes dos Estados membros do CEPEJ e da sociedade civil. O nível de resposta foi relativamente baixo e não permitiu identificar tendências claras. Alguns operadores privados não pareceram muito receptivos a esta pesquisa e os membros do CEPEJ, que pertencem em sua maior parte aos ministérios da justiça ou aos conselhos superiores de justiça, foram capazes de citar apenas as ferramentas utilizadas atualmente pela esfera pública.

14. Como resultado, o inventário abaixo é apenas parcial e baseia-se unicamente em pesquisa conduzida por especialistas e pela secretaria¹¹.

15. As classificações podem ser feitas de acordo com o serviço oferecido. O envolvimento da IA pode variar muito de acordo com as aplicações. Para fins ilustrativos, as principais categorias são as seguintes:

- Mecanismos avançados de busca de jurisprudência
- Resolução de litígios online
- Assistência na elaboração de atos

¹⁰ Veja COMPAS algoritmos ou ferramentas como RAVEL LAW ou ROSS chatbot

¹¹ Veja a bibliografia resumida no anexo IV - contribuições substanciais de Benoît Charpentier bem como Giuseppe Contissa e Giovanni Sartori (https://media.wix.com/ugd/c21db1_14b04c49ba7f46bf9a5d88581cbda172.pdf) e Emmanuel Barthe (<http://www.precisement.org/blog/Intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html#nb14>) (somente em francês)

- Análise (preditiva, escalas)
- Categorização de contratos, de acordo com diferentes critérios e detecção de cláusulas contratuais divergentes ou incompatíveis
- “Chatbots” para informar litigantes ou apoiá-los em seus procedimentos legais

16. A Letônia declarou que estava a explorar as possibilidades da aprendizagem automática para a administração da justiça. O objetivo principal seria processar judicialmente estatísticas para elaborar estimativas provisórias de recursos humanos e financeiros para ser alocado.

17. Outras atividades realizadas por empresas jurídicas de tecnologia não foram incluídos nesta classificação porque envolvem pouco ou nenhum processamento de inteligência artificial: alguns sites oferecem acesso a informações legais, soluções em nuvem, assinaturas eletrônicas, etc.

18. Segue uma lista não exaustiva de serviços jurídicos que utilizam a inteligência artificial em suas operações:

Software	País	Tipo
Doctrine.fr	França	Motor de pesquisa
Prédicite	França	Análise (exceto casos criminais)
Case Law Analytics	França	Análise (exceto casos criminais)
sData Analytics (LexisNexis)	França	Motor de pesquisa, Análise (exceto casos criminais)
Luminance	Reino Unido	Análise
Watson/Ross (IBM)	Estados Unidos	Análise
HART	Reino Unido	Análise (criminal, risco de
Lex Machina	Estados Unidos	Análise

2. Situação das políticas públicas de decisões de dados abertos nos sistemas judiciais dos Estados membros do Conselho da Europa

A disponibilidade de dados é uma condição essencial para o desenvolvimento da IA, permitindo-lhe executar determinadas tarefas, anteriormente realizadas por seres humanos de forma não automatizada. Quanto mais dados disponíveis, mais a IA consegue refinar os modelos melhorando sua capacidade preditiva. Uma abordagem de dados abertos às decisões judiciais é, portanto, um pré-requisito para o trabalho das empresas de tecnologia jurídica em pesquisa motora ou análise de tendências (“justiça preditiva”).

O processamento destes dados levanta várias questões, como a alterações na formação da jurisprudência e a proteção dos dados pessoais (incluindo os nomes dos profissionais).

19. Dados levantados pela ciência da computação são considerados o “óleo” do século XXI, já que sua exploração e cruzamento estão produzindo uma nova riqueza. Apesar de algumas partes interessadas e autores contestarem esse argumento, os sucessos globais da indústria digital nas últimas décadas confirmaram o enorme potencial de crescimento deste campo de atividade.

20. A quantificação das atividades humanas, agora em escala global, não escapam dos dados produzidos pelo setor público. Isso provocou um movimento para abrir os dados públicos, com base nos princípios fundadores do estado de direito.

21. O maior desenvolvimento dos últimos anos vem da abertura de dados públicos (dados abertos), nomeadamente no âmbito da “Partnership for Open Government” (OGP), em português “Parceria para o Governo Aberto”. A OGP é uma organização não governamental que reúne cerca de 70 Estados membros (incluindo muitos dos Estados membros do Conselho da Europa) com representantes da sociedade civil e gigantes digitais. O objetivo desta abertura é melhorar a transparência das atividades públicas, incentivar os cidadãos no desenvolvimento e avaliação de políticas públicas e garantir a integridade do serviço público, bem como dos seus agentes, pelo processamento de quantidades consideráveis de informações estruturadas em bases de dados (big data).

2.1. Definição de dados abertos sobre decisões judiciais

22. Antes de tudo, vamos redefinir a noção de dados abertos antes de lidar com a questão do impacto da abertura de dados sobre a atividade judicial. Primeiramente, muitas vezes há confusão entre o acesso à informação e o acesso aos dados (mais precisamente, o acesso a informações na forma de banco de dados)¹².

23. Certa quantidade de informação pública, exigindo ampla publicidade, já é disseminada usando tecnologia da informação. Na França, o site governamental Légifrance.fr é a principal fonte online de informação pública certificada, incluindo não só textos legislativos e regulamentares, mas também jurisprudência e informações sobre nomeações para cargos públicos. Esta informação centralizada, embora disponível na Internet, difere completamente do acesso direto aos dados organizados e incluídos em um banco de dados que pode ser baixado e processado por um computador.

24. Os dados abertos, portanto, envolvem apenas a disseminação de dados “brutos” nas bases de dados de informática estruturada. Estes dados, agregados no todo ou em parte com outras fontes estruturadas, constituem o que chamamos de big data. O Comitê Consultivo da Convenção 108 do Conselho da Europa define big data como “a capacidade tecnológica para coletar, processar e extrair conhecimentos novos e preditivos muito rapidamente a partir de um grande volume, velocidade e variedade de dados”. Em termos de proteção de dados, as principais questões não dizem apenas respeito ao volume, velocidade e variedade de dados processados, mas também a análise dos dados usando software para extrair conhecimento novo e preditivo, para fins de guiar um processo de tomada de decisão em relação a indivíduos ou grupos. Para efeitos das presentes orientações, a definição de big data abrange, portanto, tanto os dados em si quanto o processo analítico¹³.

25. Como esta definição mostra, os dados abertos não devem ser confundidos com seus meios de processamento. Muitos discursos sobre esta abertura, na verdade refere-se ao processamento realizado por vários métodos avançados

¹² Os dados são letras e números sem sentido. Informações são dados incluídos em um contexto. É o contexto que dá sentido aos dados. Podemos supor que 2005 é um ano, mas sem contexto, não sabemos. Mas em “em 2005, completamos 3 projetos” o contexto dá significado para o número. Portanto, “dados abertos” não são dados no sentido da definição, mas informação. Da mesma forma, dados grandes também são grandes quantidades de informação, não de dados.

¹³ T-PD (2017) 1, Diretrizes para a Proteção de Dados Pessoais na Era do Big Data.

que são geralmente definidos como ciência de dados. Justiça preditiva usando inteligência artificial, motores de busca avançados que aplicam critérios extremamente precisos e robôs legais, são todos aplicativos algorítmicos que são alimentados com dados, mas não têm nada a ver com a própria política de abertura em si.

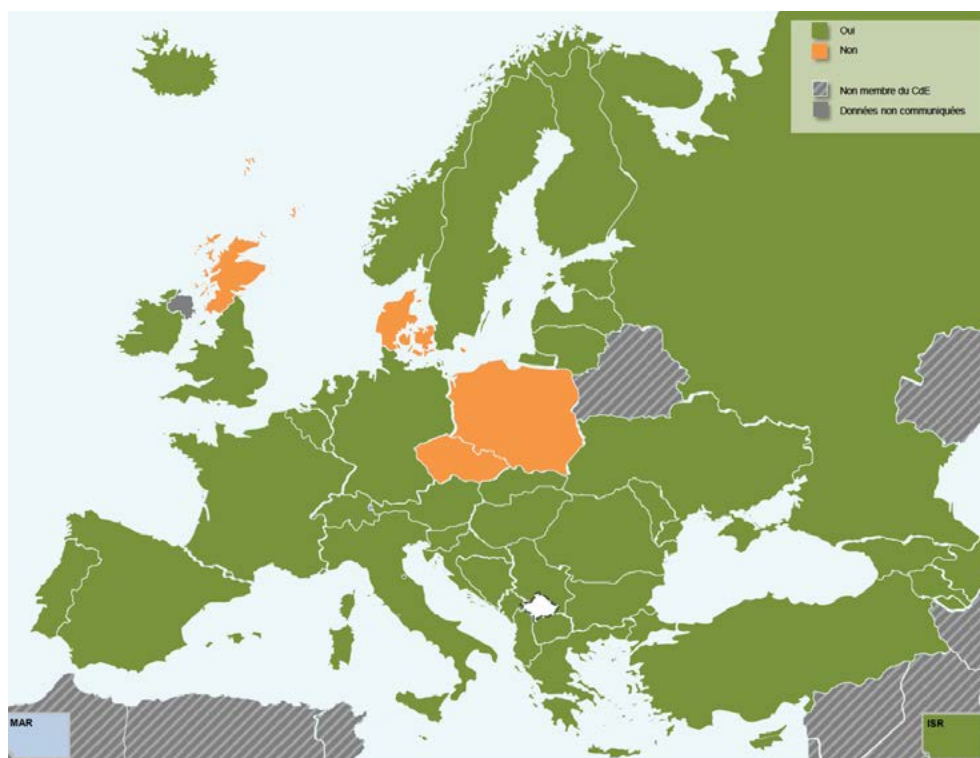
26. No entanto, esta política deve ser examinada à luz das possibilidades que oferece para processamento posterior, seja qual for sua natureza. Se determinados dados são filtrados, tendo em conta, por exemplo, a necessidade de confidencialidade e respeito à privacidade, os riscos subsequentes de uso indevido parecem ser reduzidos.

2.2. Estado de Desenvolvimento de Decisões Judiciais em Dados Abertos nos Estados membros do Conselho da Europa e as consequências para o desenvolvimento de jurisprudência

27. Qual é a situação dos Estados membros do Conselho da Europa em relação aos dados abertos sobre decisões judiciais? O ciclo de avaliação do CEPEJ 2016-2018 centrou-se pela primeira vez na questão das decisões judiciais de dados abertos, fonte de alguns tratamentos baseados em IA. A questão da anonimização ou pseudonimização dos dados no âmbito do quadro europeu de proteção de dados¹⁴ fornecido pelo Regulamento Geral de Proteção de Dados (RGPD, Regulamento UE 2016/679) e pela Convenção n.º 108 do Conselho da Europa foi objeto de uma questão específica destinada a identificar as medidas implementadas pelos Estados-Membros e observadores nesta área particularmente sensível.

¹⁴ Regulamento Geral de Proteção de Dados (RGPD, Regulamento UE 2016/679 e Convenção n.º 108 do Conselho da Europa para a proteção das pessoas no que diz respeito ao processamento automático de dados pessoais).

Figura 1. Tabela ilustrativa dos estados membros do Conselho da Europa que introduziram dados abertos das decisões judiciais em 2016 (Q62-4 do questionário de avaliação)



28. Do total de Estados e observadores avaliados, apenas 5 declararam que não haviam implementado uma política de dados abertos para decisões judiciais em 2016. Se esta taxa de resposta for relativizada, em vista da confusão de algumas respostas que igualam o acesso público às decisões com dados abertos (por exemplo, Arménia, Bélgica, Bósnia e Herzegovina, Luxemburgo, Federação Russa, Espanha, Turquia), revela, por um lado, um desejo de transparência por parte das instituições judiciais e, por outro lado, um desejo por parte de muitos países de tornar públicas as decisões dos tribunais e, assim, tornar possível negociar com eles mais tarde usando ferramentas de IA. Isso também requer esforços das instituições envolvidas, uma vez que devem ser aplicadas algumas medidas técnicas para este fim. Na França, especificamente, algumas das jurisprudências administrativas já estão disponíveis para download no site data.gouv.fr (veja abaixo).

29. No que diz respeito à proteção de dados pessoais, 23 países declararam que eles são pseudonimistas¹⁵ pelo menos em alguns tipos de contencioso (por exemplo, família), apagando dados que tornam as partes ou testemunhas identificáveis (nomes, endereços, números de telefone, números de identidade, números de contas bancárias, números de impostos, status de saúde, etc.). Este trabalho parece estar sob a responsabilidade do pessoal judicial (por exemplo, Israel, República da Moldávia) ou funcionários públicos (por exemplo, Bósnia e Herzegovina, Espanha). A Bósnia e Herzegovina e a Hungria, por outro lado, declararam publicar os nomes dos profissionais.

30. No entanto, existe uma dificuldade real em medir o impacto dos dados abertos sobre a eficiência e qualidade da justiça. Como indicado acima, a iniciativa de reutilizar esses dados é essencialmente privado, visando uma clientela profissional (advogados, departamentos jurídicos) e parece difícil identificar exaustivamente essas conquistas no contexto de apenas uma atividade intergovernamental.

31. A experiência francesa é representativa das questões levantadas por esta abordagem e revela uma série de problemas. Em primeiro lugar, é importante sublinhar que a França legislou em 2016, impondo um quadro obrigatório para a divulgação de dados abertos das decisões nos seus tribunais.

32. Os artigos 20 e 21 da Lei para uma República Digital¹⁶ romperam com a lógica anterior¹⁷ de selecionar quais decisões dos tribunais judiciais e tribunais administrativos deveriam ser disseminadas se fossem “de interesse”. Sob a nova lei francesa, no entanto, foi oposto o princípio de que tudo é publicável, exceto em casos especificado por lei (apenas para decisões judiciais) e em matéria de privacidade das pessoas envolvidas. Contudo, prevê-se que esta disponibilização ao público das decisões judiciais e administrativas é precedida de uma análise do risco de reidentificação de pessoas.

15 Conforme definido pelo T-PD em suas “Diretrizes para a proteção de indivíduos com relação ao processamento de dados pessoais em um mundo de big data” T-PD (2017) 1, pseudonimização refere-se ao processamento de dados pessoais “de tal forma que os dados pessoais não são atribuídos a um dado específico sem ter de recorrer a informações adicionais, desde que essas informações adicionais sejam mantidas separadamente e estejam sujeitas a medidas organizacionais para garantir que os dados pessoais não sejam atribuídos a pessoa singular identificada ou identificável.”

16 Esta lei foi adotada com o objetivo de alinhar a lei francesa pela Diretiva 2013/37 / UE do Parlamento Europeu e do Conselho Europeu de 26 de junho de 2013, que, por sua vez, alterou a Diretiva do Conselho, de 17 de Novembro de 2003, relativa à reutilização de informações do setor público (conhecida por “Diretiva PSI”).

17 Artigo R433-3 do Código de organização judiciária.

33. Existem inúmeras vantagens para uma ampla divulgação, promovida em especial pelo Tribunal de Cassação em duas conferências, em outubro de 2016 e fevereiro de 2018. Elas incluem uma maior conscientização da atividade judicial e tendências legais, o aumento da qualidade de um sistema de justiça que sabe o que está sendo observado e a criação de uma nova base de dados.

34. Este desejo teórico de “disciplinar” a jurisprudência através de alavancas digitais levanta, no entanto, uma série de questões gerais que são relevantes para todos os países, considerando uma abordagem semelhante. Em primeiro lugar, deve ser colocado no contexto alguns dos princípios estabelecidos pelo Tribunal Europeu dos Direitos Humanos em processos que envolvem diferenças de jurisprudência. O Tribunal enfatiza claramente a necessidade de equilibrar a segurança jurídica, o que torna as decisões mais previsíveis, com a “vitalidade” na interpretação judicial¹⁸.

35. Vários elementos devem ser examinados a este respeito, em primeiro lugar os elementos técnicos:

a. A coleta de todas as decisões judiciais elegíveis para publicação não é necessariamente bem coordenada entre todos os tribunais: algumas das aplicações comerciais nos tribunais europeus não foram concebidas para isso, particularmente em relação às decisões de primeira instância. Em alguns países é necessária a criação de novos processos para a coleta de dados de julgamentos;

b. Apesar de alguns desenvolvimentos promissores, um mecanismo automatizado de anonimização posterior capaz de prevenir qualquer risco de identificação ou reidentificação ainda não foi concebido;

c. A transformação da própria lógica de produção de jurisprudência deve ser considerada. Qual é o valor do “padrão” resultante do número de decisões tomadas sobre um assunto específico? Este “padrão” adiciona à lei? Se sim, é uma nova fonte de direito?

d. Os juízes devem ser solicitados a fornecer explicações adicionais para decisões que se desviem desta “média”?¹⁹ Esta questão não é insignificante

18 Paróquia greco-católica Lupeni e outros c. Romênia [GC]. n. 76943/11, 29/11/2016, § 116.

19 Eloi Buat-Menard and Paolo Giambiasi, «La mémoire numérique des décisions judiciaires», Dalloz Reports, 2017, p. 1483. (somente em francês)

e não encontra uma resposta uniforme em todos os sistemas judiciais. Para se ater ao direito civil francês²⁰: “O juiz decidirá a controvérsia de acordo com as regras do que lhe são aplicáveis”. No contexto francês, se fossem convidados a justificar suas decisões, fornecendo todas as razões pelas quais se desviaram da tendência da suposta maioria da jurisprudência sobre como resolver o litígio (respeitando as regras de direito relevantes), isto não seria equivalente a excluir sua autonomia? O juiz não resolveria a disputa de acordo com as regras por lei, mas também de acordo com as tendências jurisprudenciais desenvolvidas por uma ferramenta digital (que também pode ser tendenciosa ou desenvolvida sem controle externo, por um operador privado, consulte as seções 6.1 e 9 abaixo). A questão que surge vai muito além de uma característica nacional específica e se refere mais amplamente ao lugar das ferramentas de IA nos processos de tomada de decisão dos juízes. Observe que esta análise está limitada a um tratamento da jurisprudência por algoritmos e de nenhuma maneira se refere ao papel da jurisprudência nas fontes do direito ou à autoridade dos precedentes jurisprudenciais, que, além disso, são princípios bem estabelecidos a nível europeu²¹.

e. Além disso, o “padrão” do número não tenderia a nivelar as decisões da justiça, sem ordená-los de acordo com a hierarquia das jurisdições a partir das quais eles emanam, o que desrespeitaria o alcance das decisões dos tribunais supremos que em muitos Estados europeus são os garantes da interpretação uniforme da lei? Qual seria a proporção entre “a norma do número” e a jurisprudência deles? Desde que eles derivam da maioria, eles se tornariam um critério para esses tribunais quando determinar sua própria jurisprudência, que por sua vez teriam que justificar quando se desviem da visão majoritária?

f. Finalmente, não há risco de cristalização da jurisprudência, segundo uma lógica reprodutiva? Embora as decisões judiciais devam evoluir de acordo com a evolução do quadro normativo (nacional, europeu ou internacional), da referência (por exemplo, tribunais supremos, tribunais europeus) ou do critério socioeconômico, o “padrão” do número não se tornaria um padrão para

20 Artigo 12 do Código de Processo Civil francês

21 O TJUE declarou, sobre as condições que regem a responsabilidade de um Estado-Membro pelo conteúdo de uma decisão de um tribunal nacional supremo, que “uma violação da legislação será suficientemente grave” e deve dar lugar a uma indemnização pelo dano” quando a decisão em causa foi feita em violação manifesta da jurisprudência do Tribunal de Justiça sobre o assunto” (TJUE, processo C-224/01, Koebler, §56).

o qual os magistrados seriam encorajados a referir-se sem questionamentos, com um efeito induzido da excessiva padronização das decisões judiciais?

36. Por outro lado, podem-se questionar as consequências para o usuário, que supostamente é o primeiro beneficiário da transparência da ação: será que eles realmente se beneficiarão da publicação de todas as decisões judiciais na Internet ou não estarão mais propensos a se sentirem massa de decisões, sem necessariamente derivar benefícios em termos de qualidade, por falta de conhecimento fundamental para analisá-los?

37. Paradoxalmente, é questionável se os dados abertos são capazes de fornecer informações significativas, uma vez que certamente tornará possível baixar um conjunto considerável de dados brutos com um simples clique em um link, mas o significado permanecerá totalmente obscuro para a maioria dos cidadãos.

38. O primeiro requisito para que os dados abertos sejam transparentes e informativos é que possibilitem à terceiros usá-los para análise e desenvolvimento. Um modelo econômico em que os dados da jurisprudência pública, que são objetos de trabalho pelas autoridades judiciais, seriam processados gratuitamente pelo setor privado, a fim de torná-los tecnicamente “legíveis” pela IA e em conformidade com os requisitos legais para a proteção pessoal de dados, e depois vendidos aos tribunais, profissionais e cidadãos, portanto, merece ser questionado e até criticado.

2.3 A proteção de dados pessoais nas políticas de abertura de decisões judiciais

2.3.1. Os nomes das partes e testemunhas

39. A fim de alcançar um equilíbrio justo na era digital entre a necessidade de tornar públicas as decisões judiciais e respeitar os direitos fundamentais das partes ou testemunhas, os seus nomes e endereços não devem aparecer em decisões, particularmente em vista do risco de apropriação indébita e reutilização de tais informações pessoais e a sensibilidade particular dos dados que possam estar contido nas decisões²². A ocultação dessa informação pode ser feita sistematicamente através da implementação de processos automatizados.

²² Veja CEDH, Z. c. Finlandi no. 22009/93, §§ 95 e segs. E o Livro Verde sobre a informação do setor público na sociedade da informação da Comissão Europeia COM (1995) 585 (“Se não forem tomadas precauções especiais, as bases de dados jurisprudenciais, que são instrumentos de documentação jurídica, podem se tornar arquivos de informações sobre indivíduos e esses bancos de dados podem ser consultados para obter uma lista de decisões judiciais sobre um indivíduo específico, em vez de se informar sobre a jurisprudência”).

40. Outras informações “identificadoras” também podem ser obscurecidas (por exemplo, número de telefone, endereço de e-mail, data de nascimento, nomes dos filhos, apelidos e nomes de lugares). Em termos do princípio da proteção de dados pessoais, essa ocultação equivale a uma simples pseudonimização dos dados, não à anonimização completa. De fato, o volume e a variedade de informações contidas nas decisões judiciais, combinadas com a crescente facilidade de cruzar com outras bases de dados, impossibilita, na prática, garantir que a pessoa parte processo não seja não reidentificada. Na ausência de tal garantia, esses dados não podem ser qualificados como anônimos e são sujeitos, por conseguinte, às regras de proteção de dados pessoais.

41. A sensibilidade de certos dados pessoais merece atenção especial, como previsto no Artigo 6 da Convenção 108. Isso se aplica aos dados reveladores da origem étnica ou racial, opiniões políticas, filiação sindical, crenças religiosas ou outras, bem como saúde (física ou mental) ou vida sexual, que pertence à esfera mais íntima da pessoa em questão.

42. As decisões judiciais podem conter outros tipos muito variados de dados pessoais que se enquadram nesta categoria de dados sensíveis. Os tribunais que lidam com questões criminais são particularmente propensos a processar dados sensíveis, como processos e condenações. Todos esses dados sensíveis, portanto, merecem uma vigilância especial. Sua disseminação em massa apresentaria sérios riscos de discriminação, definição de perfis²³ e violação da dignidade humana.

2.3.2. Os nomes dos profissionais, incluindo juízes

43. Obviamente, saber como um julgamento será alcançado é um elemento para os advogados na previsão do resultado de um caso e eles acreditam que conhecer o juiz é quase tão importante quanto conhecer a lei. Eles vêm, há muito tempo, fazendo comparações entre juízes, de uma forma mais ou menos empírica, para melhor aconselhar seus clientes a lidar com um determinado juiz ou grupo de juízes.

44. Esse método era suficiente quando um advogado trabalhava em um número limitado de tribunais, mas o afrouxamento gradual das restrições locais em

²³ A definição de perfis é definida no Artigo 4 do GDPR. Este é um processamento usando os dados pessoais de um indivíduo a fim de analisar e prever o seu comportamento ou a sua situação, tal como determinar seu desempenho no trabalho, situação financeira, saúde, preferências, hábitos de vida, etc.

muitos países e a liberdade de se mover e trabalhar dentro da União Europeia torna razoável que qualquer advogado nacional ou europeu pretenda conhecer a jurisprudência de cada jurisdição nacional ou europeia perante a qual ele pode advogar.

45. Não podemos, portanto, excluir a possibilidade de que, no futuro, as aplicações de alto valor agregado e, portanto, muito caras, serão muito mais eficazes do que a experiência e o “bom senso” dos advogados de contencioso que trabalham com os casos de maneira tradicional. O uso de tais aplicativos poderia acentuar ainda mais a distorção da concorrência e a desigualdade de armas entre os escritórios de advocacia que usarem ou não esse software de análise “preditiva” da jurisprudência.

46. Existe um risco real de que, em benefício de tal vantagem competitiva, o princípio de um julgamento justo estabelecido por lei seja prejudicado. Com efeito, as possibilidades de traçar o “perfil do juiz”, decorrente do cruzamento de dados públicos e privados, podem permitir que empresas privadas e seus advogados se empenhem em práticas de compra de fóruns. Essa prática já foi observada por um longo tempo nos Estados Unidos e na França no litígio de infrações à imprensa e violações da privacidade por parte da imprensa, onde requerentes já são conhecidos por escolher o tribunal que ele considera conceder os maiores montantes de danos e juros.

47. Além disso, muito do discurso sobre este assunto confunde dados abertos com a necessidade de publicar certa quantidade de informação pública. Como resultado, às vezes, argumenta-se que os nomes dos profissionais devam aparecer em dados, por razões de publicidade e transparência.

48. No entanto, o fornecimento de dados informatizados jurisprudenciais é uma questão totalmente distinta dos princípios da publicação de cópias originais ou autenticadas de decisões. O objetivo dos dados abertos é permitir o processamento automatizado de jurisprudência, sob uma licença de exploração de baixo custo. Como indicado anteriormente, esta disposição é feita sob a forma de uma base de dados abrangente, opaca e não diretamente compreensível pelos cidadãos.

49. Esta disposição obviamente não atende à necessidade de divulgar os nomes de profissionais que contribuíram para uma decisão específica. Note-se que,

na lei, a menção dos nomes dos juizes em decisões judiciais é uma obrigação comum dos Estados-Membros, em conexão com o princípio de publicidade do artigo 6º, nº 1 da Convenção Europeia dos Direitos Humanos, a fim de garantir a imparcialidade objetiva dos juizes (que devem ser identificáveis, legalmente nomeados e designados para as funções que desempenham) e o cumprimento das regras processuais (por exemplo, publicidade e colegialidade)²⁴.

50. A resposta à questão da legitimidade ou não da publicação dos nomes de profissionais²⁵ em dados abertos, portanto, não tem nada a ver com a obrigação de publicar os nomes dos profissionais nas decisões. Pelo contrário, parece que o desafio está em conciliar os requisitos muitas vezes conflitantes: de um lado, tornar as atividades públicas transparentes, permitindo que os cidadãos conheçam e avaliem seus juizes, do outro lado, proteger a privacidade dos profissionais (cujas funções não devem limitar suas garantias fundamentais neste campo). Existem desafios rigorosos no que diz respeito à garantia da imparcialidade dos juizes e até das instituições judiciais como um todo, acerca de quais políticas de dados abertos são realmente projetadas para fortalecê-la²⁶. Que medidas práticas podem ser tomadas para protegê-los de potenciais tentativas de desestabilização que derivam da referência cruzada dos dados nominativos dos juizes, presentes na base de dados de outras fontes (redes sociais, sites comerciais), com o objetivo de identificar vieses políticos, religiosos e outros hipotéticos?

51. Estas questões não surgem da mesma forma em toda a Europa e dependem das características específicas do sistema judicial em questão (e da natureza do órgão de administração da carreira do judiciário), da natureza colegial ou não do julgamento e do nível da autoridade envolvida. Na Suíça, por exemplo, quando os juizes são eleitos, a publicação é uma garantia de transparência e responsabilidade social dos juizes em relação aos cidadãos, bem como aos grupos políticos. Esta informação já está disponível nos motores de busca

24 CEDH, *Vernes c. França*, n.º 30183/06, no que diz respeito à identificação dos juizes que proferiram a decisão, *Pretto v. Itália*, N.º 7984/77 para a publicação do julgamento, - *Kontalexis v. Grécia*, n.º 59000/08, § 38, *DMD GROUP, a.s., v. Eslováquia*, n.º 19334/03, § 66, *Miracle Europe KFT v. Hungria*, n.º 57774/13, § 58 para o direito a um tribunal estabelecido por lei em conjunto com o direito a um juiz imparcial.

25 Estas questões também podem ser aplicadas ao funcionário que toma decisões judiciais e aos funcionários assistentes do juiz mencionados na composição da formação do juízo (ainda que em menor grau).

26 Ver ECHR *Previti v. Itália*, N.º 45291/06, §§ 249 e seguintes, que relembra os princípios da imparcialidade do juiz

online (que, estritamente falando, não são dados abertos)²⁷.

52. Estas questões também não surgem da mesma forma, dependendo do nível de jurisdição. O valor da caracterização da jurisprudência dos juízes de primeira instância pode não ser o mesmo que para os juízes dos tribunais supremos ou dos tribunais internacionais. Por exemplo, o Tribunal Europeu dos Direitos Humanos autoriza a pesquisa de julgamentos pelos nomes dos juízes membros do painel de decisão, mas não permite o cálculo de estatísticas relativas a um determinado juiz²⁸. Por outro lado, em países onde os órgãos judiciais não estão familiarizados com a prática de opiniões divergentes (existentes neste tribunal internacional), pode parecer injusto atribuir a um juiz responsabilidade pessoal por uma decisão que ele votou contra, durante a deliberação em um tribunal colegiado.

53. Estes debates foram bem definidos por uma missão de estudo realizada na França pelo professor Loïc Cadiet. As conclusões da missão permanecem limitadas, pois não recomendam proibir a publicação, mas reservar a publicação para certos tipos de litígios e excluí-las em outros (por exemplo, nas questões criminais especializadas). A possibilidade de publicar apenas os nomes dos juízes do Supremo foi proposta, embora tenha sido admitido que isto pode resultar em uma “viagem sem volta”.

54. Por conseguinte, pode ser aplicado um simples princípio de precaução, tal como é previsto, para avaliar o interesse, por tipo de litígio e grau de jurisdição, na publicação dos nomes dos profissionais em um banco de dados para download. Da mesma forma, não se pode excluir a possibilidade de as próprias instituições judiciais ou terceiros autorizados explorarem essas informações fora do contexto de dados abertos para descobrir a jurisprudência.

55. De qualquer forma, a publicação deste tipo ainda terá de ser examinada à luz das regras europeias internacionais aplicáveis à proteção de dados pessoais, como as da Convenção Nº 108 e da União Europeia (GDPR e Diretiva 680/2016).

²⁷ Veja o exemplo do Tribunal Federal Suíço, cuja jurisprudência está disponível para download: (<https://www.bger.ch/en/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-recht.htm>) ou cantões (cantão de Genebra por exemplo <http://ge.ch/justice/in-jurisprudence>).

²⁸ As decisões do Tribunal Europeu dos Direitos Humanos são colegiais. As publicações incluem quaisquer opiniões divergentes.

3. Características operacionais da inteligência artificial (aprendizado de máquina) aplicada a decisão judicial

Processamento de linguagem natural e aprendizado de máquina são as duas principais técnicas de processamento de decisões judiciais usando inteligência artificial.

Na maioria dos casos, o objetivo desses sistemas não é reproduzir o raciocínio jurídico, mas identificar as correlações entre os diferentes parâmetros de uma decisão (por exemplo, em um pedido de divórcio, a duração do casamento, a renda dos cônjuges, a existência de adultério, o valor do benefício declarado, etc.) e, através do uso de aprendizado de máquina, deduzir um ou mais modelos. Tais modelos serão utilizados para “prever” ou “antecipar” uma futura decisão judicial.

O objetivo deste capítulo é esclarecer a natureza intrínseca do software descrito como “preditivo” - às vezes versões “beta”, ou seja, em fase de desenvolvimento - tanto em termos do seu potencial como das suas limitações. Pretende-se fornecer uma explicação simples dos algoritmos de aprendizagem de máquina, que estão no centro da análise automatizada da jurisprudência.

3.1 Funcionalidades teóricas do chamado software de justiça preditiva

56. Como introdução, é necessário revisar brevemente os recursos prometidos pelo software “preditivo”. Eles propõem estabelecer probabilidades de sucesso (fracasso) de um processo perante um tribunal. Essas probabilidades são estabelecidas através da modelagem estatística de decisões anteriores usando métodos a partir de dois grandes domínios da ciência da computação: processamento de linguagem natural e aprendizado de máquina. Essas abordagens de modelagem estatística são frequentemente chamadas de inteligência artificial (IA); na verdade, trata-se de IAs “fracas” (veja o glossário)

57. Deve ficar claro de imediato que o termo Inteligência Artificial (IA) é discutido por especialistas, pois leva a muitas ambiguidades. O termo IA entrou agora em nossa linguagem cotidiana para descrever uma gama diversificada de ciências e tecnologias que permitem aos computadores vencer os melhores campeões no jogo de Go²⁹, dirigir um carro, conversar com humanos, etc. Pesquisadores preferem identificar as diferentes aplicações através das tecnologias exatas

²⁹ <https://www.nature.com/articles/nature16961>

que as sustentam, incluindo o aprendizado de máquina, às vezes referindo-se a todos esses recursos de IA altamente especializados como IAs “fracas” (ou “moderadas”). Estes são distinguidos de um objetivo final - ainda totalmente teórico - de criar um IA “forte”, ou seja, uma máquina de autoaprendizagem capaz de compreender automaticamente o mundo em geral, em toda a sua complexidade.

58. Especificamente para a justiça, os chamados sistemas de justiça preditiva visam ser usados por departamentos jurídicos, seguradoras (tanto para suas necessidades internas como para seus segurados), bem como advogados para antecipar o resultado do litígio. Eles também poderiam ser utilizados para auxiliar os magistrados na tomada de decisões.

59. Eles fornecem uma representação gráfica das probabilidades de sucesso para o resultado contencioso, de acordo com critérios inseridos pelo usuário (específico para cada tipo de contencioso). Estes sistemas afirmam ser capazes de calcular as quantias prováveis da compensação distribuída pelos tribunais.

3.2. O funcionamento concreto da inteligência artificial: máquinas estatísticas construindo modelos baseados no passado

60. Uma distinção deve ser feita desde o início entre o que é uma “predição” e o que é uma “previsão”. A previsão é o ato de anunciar o que irá acontecer (antes, digamos) em eventos futuros (por inspiração sobrenatural, por clarividência ou premonição). Predição, por outro lado, é o resultado da observação (apontando, vendo) um conjunto de dados, a fim de prever uma situação futura. Este abuso de linguagem e sua disseminação parece ser explicado por uma transferência do termo das ciências “duras”, onde refere-se a uma variedade de técnicas de ciência de dados derivadas da matemática, estatística e teoria de jogos que analisam fatos presentes e passados para fazer hipóteses sobre o conteúdo de eventos futuros.

61. Também deve ser notado que a lógica da operação de um software de justiça preditiva depende principalmente de métodos generativos (comumente conhecidos como Bayesianos) ou abordagens discriminatórias que tentam estimar o intervalo atual ou futuro de valores para uma variável (por exemplo, o resultado de um teste) da análise de exemplos anteriores.

62. De um modo geral, também é importante ter em mente a noção antropomórfica de que as máquinas computacionais são inteligentes e que seus projetistas conseguiram inserir uma mente dentro de seus mecanismos³⁰. Infelizmente, essa ideia ainda permeia muitas análises de justiça preditiva que emprestam a esses dispositivos capacidades imediatas ou futuras para a quase replicação da inteligência humana. Esse contexto, alimentado todos os dias por uma série de avanços revolucionários da IA³¹, nos leva a abordar essas ferramentas com aspectos preditivos com uma certa dose, consciente ou não, de misticismo, por vezes afirmando que o que não é totalmente possível hoje será inevitavelmente possível amanhã.

63. Os “fortes” IAs da literatura de ficção científica não existem. Esse tipo de IA, que estaria equipada não apenas com inteligência, mas também com consciência, permanece puramente ficcional. Os sistemas de aprendizado de máquina atualmente em desenvolvimento são descritos como IAs “fracos” e são capazes de extrair padrões complexos e aprender com grandes volumes de dados de forma eficiente e muitas vezes com altos níveis de precisão preditiva.

64. Para superar quaisquer considerações instintivas ou fantasiosas, essas tecnologias de processamento e análise de informações devem ser entendidas com base nos três conceitos seguintes:

- IA não é um objeto único e homogêneo: é na verdade um conjunto de ciências e técnicas (matemática, estatística e ciência da computação) capaz de processar dados para projetar tarefas de processamento de computador muito complexas;

- O motor da IA não produz inteligência em si, mas funciona através de abordagem indutiva: a ideia é associar, de forma quase automatizada, um conjunto de observações (entradas) com um conjunto de resultados possíveis (saídas) usando várias propriedades pré-configuradas. Especificamente para a justiça preditiva, o motor constrói ligações entre os diferentes grupos lexicais que compõem decisões judiciais. Esses grupos são correlacionados entre aqueles identificados no estágio de entrada (fatos e raciocínio) e aqueles no

30 Dominique Cardon, A quoi rêvent les algorithmes, nos vies à l'heure des big data, La République des idées, Editions du Seuil, 2015, p. 58.

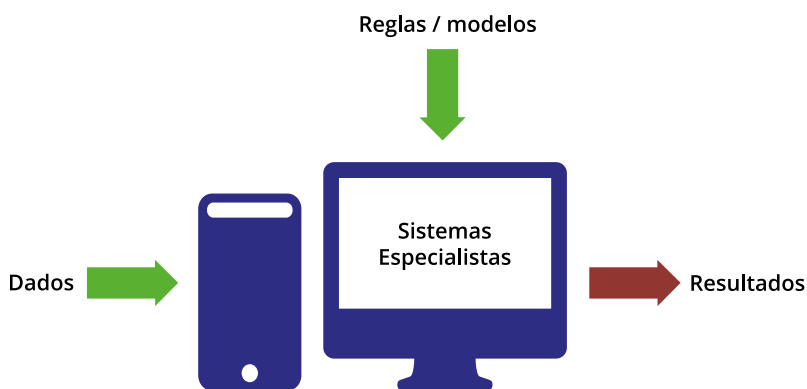
31 AlphaZero : l'IA de DeepMind apprend et maîtrise le jeu d'échecs en moins de 4 heures”, Génération NT, article published on 8 December 2017, [Online], <https://www.generationnt.com/alphazero-ia-deepmind-maitrise-echecs-4-heures-actualite-1948919.html> (página acessada em 14 de dezembro de 2017).

estágio de saída (o dispositivo da decisão) e, em seguida, classificados.

- A confiabilidade do modelo (ou função) construído depende fortemente da qualidade dos dados utilizados e a escolha da técnica de aprendizado de máquina.

65. É necessário retornar brevemente aos ancestrais desses sistemas, os sistemas especialistas (SE), que dependiam de regras de processamento escritas por um cientista da computação. Os sistemas especialistas (ES) desenvolveram-se rapidamente no final dos anos 80 e 90, especialmente nos campos da medicina e finanças³². Esses sistemas foram capazes de responder perguntas especializadas e raciocínio baseado em fatos conhecidos, executando regras de codificação predefinidas em um “motor”. Apesar do sucesso dos SEs, como o Deep Blue contra Garry Kasparov durante uma série de jogos de xadrez em 1997, estes sistemas acabaram falhando, notadamente porque eram incapazes de interpretar “a infinita variedade de situações e contextos”³³ e tornaram-se ineficazes para além de 200 a 300 regras de codificação, tanto em termos de desempenho de execução como de manutenção (o raciocínio seguido pelo sistema tornou-se quase impossível de ser apreendido por seus designers).

Fig. 2: Os antigos sistemas especialistas foram programados com regras, reproduzindo a lógica do raciocínio jurídico

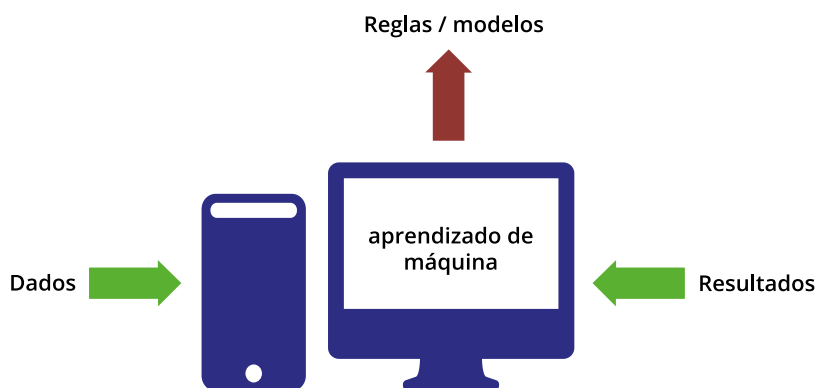


32 Um exemplo é o High Frequency Trading (HFT), em português: transação de alta frequência, que é um tipo de transação financeira realizada em alta velocidade por software baseado em algoritmos. Em 2013, na Europa, 35% das transações já foram realizadas com a HFT. Nos EUA, a HFT representa 70% do volume de negócios no mercado de ações. Em vista da automação das transações, a queda ainda é mais rápida em caso de acidentes, como foi o caso durante a crise financeira de 2007

33 Hubert Dreyfus, *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, The MIT Press, 1992 citado por Dominique Cardon em “A quoi révent les algorithmes, nos vies à l’heure des Big Data”, p.59.

66. Hoje, a ideia não é mais escrever regras de espelhamento de raciocínio, como nos sistemas especialistas mais antigos³⁴, mas permitir que os próprios sistemas de aprendizado de máquina identifiquem modelos estatísticos existentes nos dados e façam a correspondência para resultados específicos.

Fig. 3: O aprendizado de máquina sozinho produz modelos pesquisando automaticamente correlações entre dados e resultados.



67. Ao fazê-lo, estes sistemas não reproduzem ou modelam o raciocínio. Por exemplo, os tradutores on-line de hoje não carregam raciocínio abstrato. Eles inferem uma estimativa provável da melhor correspondência entre grupos de estruturas lexicais e traduções já feitas. As ações dos usuários contribuem obviamente para o aprimoramento da busca por correspondência, mas o algoritmo de aprendizado de máquina não realiza, estritamente falando, uma tradução compreendendo o significado das sentenças processadas³⁵.

68. Este exemplo mostra como a abordagem de advogados e pesquisadores podem ser diferente: um advogado procurará compreender a eficácia de regras pré-existentes, que o pesquisador só pode fornecer através da linguagem de sua ciência, e isso às vezes se limita a interpretar milhares de parâmetros de um sistema de aprendizado de máquina.

34 As regras de processamento escritas podem ser baseadas em código específico, em diferentes linguagens de programação, como LISP ou editores, para modelar regras. Veja, por exemplo, Radboud Winkels, “CLIME : Un projet de développement de serveurs juridiques intelligents”, in Danièle Bourcier, Patricia Hasset and Christophe Roquilly (eds.), Droit et intelligence artificielle, 2000, Romillat, p. 59.

35 Li Gong, “La traduction automatique statistique, comment ça marche ?”, Interstices.info, publicado em 29 de outubro de 2013, [Online], https://interstices.info/jcms/nn_72253/la-traduction-automatique-statistique-comment-ca-marche (página acessada em 14 de dezembro de 2017).

69. O aprendizado de máquina é um campo da ciência da computação no qual os programas aprendem com a experiência. Algoritmos fazem uma máquina executar um processo de treinamento, como uma criança aprendendo em seu ambiente. Em resumo, estas técnicas de aprendizagem podem ou não ser supervisionadas por um ser humano. A categoria de aprendizado de máquina mais proeminente atualmente é a da aprendizagem não supervisionada por reforço: a máquina sozinha obtém “recompensas” virtuais se a ação produzir o resultado esperado em um determinado ambiente. Métodos de aprendizado de máquina incluem redes neurais (ou sua versão mais complexa conhecida como deep learning - em português: aprendizagem profunda), que foram amplamente divulgadas em vista de sua autonomia e de suas aplicações bastante impressionantes, como a que consegue obter altas pontuações nos antigos videogames Atari 2600³⁶ (com base unicamente no posicionamento de pixels e pontuações). Outras aplicações práticas para essas tecnologias já estão afetando nossa vida cotidiana e estão começando a aparecer no mundo profissional da justiça³⁷.

70. No entanto, estes algoritmos permanecem altamente especializados para uma tarefa e apresentam problemas de discernimento diante de situações caóticas ou com dados insuficientes para permitir a previsão (como a compreensão real da linguagem natural). No campo das ciências sociais, ao qual pertencem o direito e a justiça, o fracasso pareceria até mesmo inevitável na ausência de um modelo convincente de cognição. Para Pierre Lévy, a inteligência artificial, de fato, fornece uma caixa de ferramentas com conteúdo heterogêneo (regras lógicas, sintaxes formais, métodos estatísticos, simulações neurais ou sociobiológicas, etc.) que não oferecem uma solução geral para o problema da modelagem matemática da cognição humana³⁸. Assim, a aprendizagem preditiva real deveria, na realidade, basear-se numa boa representação sistêmica do mundo, Yann LeCun, pesquisador no campo da inteligência artificial, acredita que é uma questão científica e matemática fundamental, não uma questão de tecnologia³⁹.

³⁶ <https://www.nature.com/articles/nature14236>

³⁷ IBM Watson inteligência artificial equipando o campo médico ou servindo como um motor “Ross”, apresentado como um advogado “virtual” - Ross, “Faça mais do que humanamente possível”, [Online], <http://rossintelligence.com> (acessado em 14 de dezembro de 2017).

³⁸ Pierre Lévy, “Intelligence artificielle et sciences humaines”, blog de Pierre Levy, 8 dw outubro de 2014, [on line], <https://pierrelevyblog.com/2014/10/08/intelligence-artificielle-et-sciences-humaines/> (página consultada em 30 de dezembro de 2017).

³⁹ Yann LeCun, “Qu’est-ce que l’intelligence artificielle”, pesquisa publicada no site do Collège de France, [Online], https://www.college-de-france.fr/media/yann-lecun/UPL4485925235409209505_Intelligence_Artificielle_Y_LeCun.pdf (página consultada em 14 de dezembro 2017).

71. Além disso, a singularidade dos atuais sistemas de processamento de big data não é fingir reproduzir o nosso modelo de cognição, mas produzirem estatísticas de contexto sobre um tamanho de dados sem precedentes, sem qualquer garantia real de exclusão das correlações falsas⁴⁰.

72. Embora seja comumente aceito que modelos estatísticos e probabilísticos são melhorados aumentando os dados que os alimentam, alguns matemáticos advertiram sobre os riscos de aumentar falsas correlações (isto é, ligações entre fatores que não têm absolutamente nenhum nexo de causalidade) em big data. Os matemáticos Cristian Sorin Calude e Giuseppe Longo apontam para o risco de um dilúvio de correlações falsas em big data: “quanto maior o banco de dados usado para correlações, maiores são as chances de encontrar padrões recorrentes e maior é a chance de cometer erros”.⁴¹ O que pode aparecer como regularidades para uma IA (ligações recorrentes entre diferentes dados, conceitos, contextos ou grupos lexicais), pode na verdade ser aleatório. Mesmo que o argumento dos dois matemáticos não possa ser generalizado com muita pressa, eles acreditam que em alguns grandes conjuntos de números, pontos ou objetos, aparecem padrões aleatórios regulares e parece impossível distingui-los algorítmicamente de padrões que revelam causalidades⁴².

73. Por fim, os algoritmos, independentemente do seu nível atual de sofisticação, ainda se resumem automaticamente às interações estabelecidas na fase de aprendizagem e daí para o seu passado. O conteúdo e a qualidade dos fluxos de dados utilizados na composição dos cálculos é, portanto, fundamental para a compreensão dos resultados obtidos e a identificação de possíveis vieses analíticos. Existem muitos desafios, uma vez que, numa abordagem exaustiva, a análise da maior quantidade de dados relativos a uma atividade produzirá resultados cujo significado para ser esclarecido terá que necessariamente ser informado à luz de todos os fatores que tiveram influência. Com uma abordagem mais restritiva, por amostragem de dados de entrada, haverá também riscos em vista do viés de arbitragem necessário para selecionar um dado em detrimento de outro.

40 Dominique Cardon, op. cit., p.60.

41 Cristian Sorin Calude, Giuseppe Longo, “Le déluge des corrélations fallacieuses dans le big data”, in Bernard Stiegler (dir.) (2017), La toile que nous voulons, FYP editions, p.156.

42 Teoria de “Ramsey”, Ronald L. Graham, Joel H. Spencer – “Ramsey Theory”, Scientific American, vol.263, Nº.1, Julho de 1990, p112-117 citado por Cristian Sorin Calude, Giuseppe Longo, “Le déluge des corrélations fallacieuses dans le big data”, ibid.

4. A inteligência artificial pode modelar o raciocínio jurídico?

Inteligências artificiais contornam as dificuldades encontradas com os antigos sistemas especialistas: eles não tentam espelhar manualmente o raciocínio jurídico, cuja reprodução não é em si um objetivo para eles. Aprendizado de máquina leva a categorizações entre os diferentes parâmetros identificados pelos designers ou aqueles descobertos pela máquina.

74. Os sistemas jurídicos continentais estão muito longe do ideal de racionalidade que deveria para incorporar, por exemplo, o código civil de 1804 na França. Há uma infinidade de fontes que não se encaixam perfeitamente e que se relacionam com um conjunto de regras cujo significado permanece indeterminado, o que o teórico do direito Herbert L. A. Hart chamou de “textura aberta da lei”.⁴³ Sistemas de direito comum, embora considerados mais economicamente eficientes porque se adaptam às novas necessidades legais⁴⁴, também estão evoluindo e não oferecem mais segurança jurídica. As regras legais não evoluem de forma linear, distinguindo-se das leis empíricas (aquelas das “ciências exatas”), onde cada nova regra geralmente complementa as anteriores e não invalida um conjunto completo de raciocínio.

75. Deve recordar-se que, face a esta complexidade geral do sistema jurídico, os antigos sistemas especialistas em TI rapidamente atingiram seus limites com 300 regras lógicas. A divisão da lei em “regras de produção” não foi suficiente para fornecer uma representação válida do corpo de conhecimentos e métodos que orientam um advogado.

76. Consequentemente, demonstrou-se ser igualmente impossível modelar o raciocínio dos juizes num computador, uma vez que este tem de modelar um sistema legal positivo. Como teoria do direito, o raciocínio judicial é acima de tudo uma questão de avaliação e interpretação, dos fatos comprovados e relevantes de um caso, de regras de direito aplicáveis (textuais ou precedentes), das quais o significado permanece, como já dito, indeterminado⁴⁵ e da interpretação subjetiva pelos juizes do conceito de equidade, que deve sofrer novas mudanças na Europa, com a revisão do “controle de proporcionalidade” incentivado pelo Tribunal Europeu dos Direitos Humanos.

43 Herbert L. A. Hart (1976), *Le concept de droit*, Saint-Louis university departments, Bruxelles.

44 <https://www.contrepoints.org/2014/08/15/177160-common-law-contre-droit-civillexperience-francaise-de-lancien-regime>

45 Sobre estas questões, Michel Troper (2001), *La théorie du droit, le droit, l'Etat*, PUF, Paris, spec. p. 69-84.

77. Xavier Linant de Bellefonds assinalou que “a complexidade da lei está em seu caráter teleológico e contencioso: dois raciocínios coerentes podem levar a julgamentos diferentes de acordo com duas prioridades diferentes”.⁴⁶

78. Isso porque o famoso silogismo legal é mais uma forma de apresentar o raciocínio jurídico do que a sua tradução formal. Não reflete o raciocínio completo do juiz, que é, de fato, composto por uma multiplicidade de fatores de tomada de decisão, não pode ser formalizado a priori, e às vezes é baseado em seu entendimento: quais são os fatos relevantes? Esses fatos são comprovados? Qual regra aplica-se a eles? Qual é o significado desta regra em relação ao caso a ser decidido? Qual fonte deve prevalecer entre uma variedade de fontes conflitantes? A coerência geral das decisões judiciais nunca é assegurada, porque elas têm mais a função de convencer acerca da validade de uma solução específica do que descrever estritamente e objetivamente todas as etapas que resultou na tomada da decisão.

79. No entanto, este trabalho de interpretação é precisamente o que as técnicas de aprendizagem de máquina não fazem, já que carregam processamento automatizado, onde a correlação de vasta quantidade de informação deve substituir a compreensão da real causalidade de uma decisão. Eles não tentam formalizar o raciocínio legal, mas esperam que os modelos capturados por eles possam antecipar as prováveis decisões de um juiz em situações semelhantes.

80. Os resultados alcançados pelas IAs não são, na realidade, relacionados à questão da conformidade legal de uma determinada solução e não pode discriminar entre argumentos legais e ilegais.

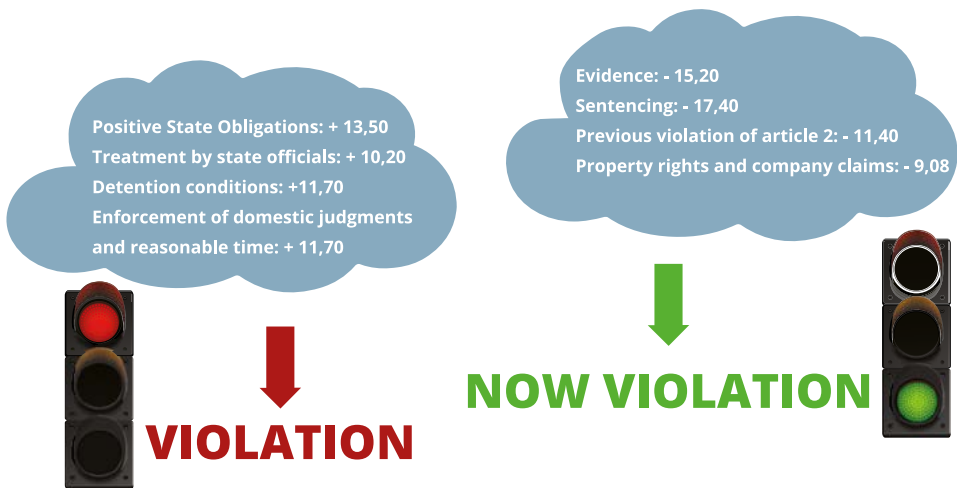
81. Uma revisão do trabalho da University College of London (UCL) sobre a jurisprudência do Tribunal Europeu dos Direitos Humanos confirma este diagnóstico. O estudo da UCL assumiu que um simples modelo de aprendizagem automática poderia prever o resultado de um caso com 79% de precisão para esse tribunal específico. A máquina modelo de aprendizagem mostrou-se mais efetiva na parte descritiva dos fatos das decisões estudadas do que da fundamentação relativa à aplicação da Convenção ao caso⁴⁷.

⁴⁶ Xavier Linant de Bellefonds (1994), “L'utilisation des systèmes experts en droit comparé”, *Revue internationale de droit comparé*, Vol. 46, N° 2, p. 703-718.

⁴⁷ Trabalhar com uma amostra de 584 decisões do Tribunal Europeu dos Direitos Humanos: Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro, Vasileios Lampos, “Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective”, publicado em 24 de outubro de 2016, [Online], <https://peerj.com/articles/cs-93/> (página acessada em 14 de dezembro de 2017).

82. O exame dos termos correlacionados com a descoberta de uma violação (com ponderação positiva) e aqueles correlacionados com a não violação da Convenção Europeia sobre Direitos Humanos (com ponderação negativa) ilumina inequivocamente a mecânica do trabalho, que não é de forma alguma comparável ao raciocínio jurídico.

Fig. 4: Ilustração do trabalho da UCL - Peso teórico atribuído a palavras ou termos de acordo com a sua ligação com as conclusões de violação ou não violação da Convenção União Europeia dos Direitos Humanos.



83. A escolha dos fatos relevantes e sua interpretação constituem um dos elementos da decisão do juiz. Em outras palavras, o estudo da UCL não foi realmente capaz de produzir uma probabilidade com o material em grande parte derivada de raciocínio e motivação do juiz e não com a da recorrente, que reuniu com base em aparência. Sua IA, portanto, estabeleceu uma alta probabilidade de correspondência entre grupos de palavras e uma decisão já formalizada. Não pode de forma alguma reproduzir o raciocínio dos juízes europeus e, acima de tudo, a previsão de um resultado, com base, por exemplo, numa conta em bruto de um futuro requerente perante o Tribunal de Estrasburgo, cuja candidatura será submetida a um exame de admissibilidade muito estrito (quase 70.356 pedidos

foram declarados inadmissíveis ou removido da lista em 2017),⁴⁸ com base em grande parte na aplicação de normas de avaliação (importância e gravidade da queixa, etc.), proporcionando uma considerável margem de manobra na tomada de decisões.

48 Fonte: 2017 Análise estatística da CEDH, janeiro de 2018

5. As IA's podem explicar o comportamento dos juízes a posteriori?

Uma explicação a posteriori do comportamento de um juiz, em particular a revelação do viés, exigiria que todos os fatores potencialmente causadores fossem identificados por meio de um quadro interpretativo e uma análise contextualizada. O fato de que, estatisticamente, a guarda de crianças é mais frequentemente confiada às mães do que aos pais não demonstra um viés por parte dos juízes, mas revela a necessidade de mobilizar diferentes disciplinas das ciências sociais para lançar luz sobre esse fenômeno.

84. Do ponto de vista científico, explicar um fenômeno ou, no que nos diz respeito, um comportamento humano, é determinar os mecanismos causais que levaram a este comportamento usando certa quantidade de dados contextuais.

85. Isto exige, de uma forma muito esquemática, a constituição preliminar de um quadro interpretativo, ele próprio derivado da observação repetida deste tipo de evento ou comportamento na presença de certos fatores ou elementos. O quadro interpretativo é constituído pelas hipóteses ou pontos de vista adotados pelas diferentes disciplinas das ciências sociais. Este é um passo analítico adicional que pode ser inserido em algoritmos, mas que eles não podem executar sozinhos.

86. Algumas empresas de tecnologia jurídica foram além e acharam que poderiam identificar possíveis vieses pessoais de juízes e alimentar suspeitas de preconceito. Inserindo os nomes dos juízes presidentes dos tribunais administrativos e da administrativa de recurso da França, construíram um indicador da taxa de rejeição de recursos contra as obrigações de deixar o território francês tomado pelas autoridades administrativas. Alguns comentaristas então apoiaram com grande entusiasmo que a chamada imparcialidade dos juízes foi “solapada pela inteligência artificial”.

87. Mas tais interpretações podem realmente ser alcançadas com base em um processamento algorítmico de decisões judiciais? Caracterizar viés pessoal dos juízes no processo de tomada de decisão consiste em fornecer uma explicação de que, neste caso sua decisão, foi determinada por seus traços de personalidade, opiniões ou religião. No entanto, como já foi dito, tal explicação causal não pode simplesmente ser deduzida do resultado probabilístico fornecido pelos algoritmos. Pelo contrário, requer trabalho analítico adicional para isolar, entre os muitos fatores correlacionados (incluindo a identidade dos membros da

formação do julgamento), aqueles que são verdadeiramente causativos. Por exemplo, o fato de um tribunal de família decidir estatisticamente que as crianças devem morar com a mãe não necessariamente reflete o preconceito do juiz em favor das mulheres, mas a existência de fatores psicossociais, econômicos e mesmo culturais específicos da jurisdição, como o tempo de trabalho de cada um dos pais, sua renda, a disponibilidade local de cuidados infantis coletivos, se a criança está ou não na escola, se ou não um dos pais está em um novo relacionamento ou mesmo simplesmente a falta de interesse de um dos pais em cuidar de uma criança pequena.

88. Do mesmo modo, as decisões sobre a expulsão de um país por um tribunal administrativo localizado perto de um grande centro de detenção não podem ser comparadas de forma justa com as de um tribunal que lida com essas disputas apenas ocasionalmente.

89. Além disso, independentemente da localização do tribunal, a questão da jurisprudência de um único juiz que só ocasionalmente lida com um certo tipo de litígio, mas que usa (ou ignora) a jurisprudência de seus colegas, é particularmente interessante e pode legitimamente levantar a questão da igualdade dos cidadãos nos Processos judiciais. No entanto, o foco deve permanecer na remodelação ou preservação da natureza colegial do sistema judicial, em vez de classificação ou estigmatização através de ferramentas de aprendizagem de máquina.

90. O que pode ser deduzido da personalidade do presidente de uma turma de julgamento colegiado, quando seu nome é a única informação nominativa visível em decisões judiciais administrativas abertas?

91. Além disso, como podemos comparar duas abordagens filosóficas e culturais diferentes para decisão judicial, quando, em alguns países europeus, incluindo a França, há uma cultura de precedente e um conhecimento detalhado pelos juízes dos bancos de dados factuais de todas as decisões de primeira e segunda instância (banco de dados Ariane) no campo da justiça administrativa, enquanto em outros países ou sistemas predomina a independência intelectual de cada tribunal, bem como a disposição de tratar cada situação caso a caso?

92. Fornecer uma explicação precisa acerca de uma decisão judicial requer, portanto, uma análise muito mais detalhada dos dados contingentes em cada

caso e regras de direito aplicáveis, em vez de sustentar uma esperança vã de que a massa de links farão sentido.

6. Como a IA deve ser aplicada na justiça civil, comercial e administrativa?

O estado de desenvolvimento das técnicas de aprendizado de máquina não torna possível hoje alcançar resultados confiáveis na “previsão” de decisões judiciais. No entanto, a sua aplicação no domínio da justiça civil, comercial e administrativa deve ser considerada para o estabelecimento de escalas ou resolução pré-contenciosa dos litígios on-line, uma vez que um recurso posterior ao juiz continua a ser possível.

93. Como podemos ver, a primeira questão levantada por tal uso da inteligência artificial não é tanto se ela é benéfica ou prejudicial, desejável ou não, mas se os algoritmos propostos podem alcançar o tipo de resultado desejado. As conclusões do experimento realizado nos tribunais de recurso de Douai e Rennes, na França, demonstram claramente que, na presença de um discurso instigante, que promove um produto da IA, pode esconder defeitos redibitórios e resultados de análise totalmente errôneos.

94. Independentemente da qualidade do software testado, a antecipação da decisão do juiz em matéria civil, comercial e administrativa parece ser uma questão potencialmente desejável, embora, por vezes, por razões muito diferentes, pelos responsáveis pelas políticas públicas judiciais e pelos profissionais do direito privado.

95. Seja qual for a tradição jurídica do país, a incerteza jurídica, ou seja, o risco de ter uma reivindicação legal validada ou rejeitada, solicita o desejo de ser capaz de quantificar esses fatores usando essas novas aplicações tecnológicas.

96. Advogados (ou departamento jurídico de uma empresa) veem a possibilidade de usar esta tecnologia para fornecer aos seus clientes melhor assessoria baseada em avaliação empírica e sistemática das chances de sucesso de um procedimento, bem como incentivar a conclusão de transações que, se necessário, possibilitarão evitar um julgamento longo e caro. Algumas seguradoras já oferecem o uso de sistemas preditivos para seus clientes avaliarem seus negócios.⁴⁹

97. Ao mesmo tempo, os julgadores consideram que esta é uma oportunidade para melhor regulamentar o fluxo de novos processos através dos tribunais e

⁴⁹ O software Preditivo é oferecido aos segurados da Allianz, por exemplo.

para reduzir os custos operacionais judiciais. Acreditam que isso incentivaria os litigantes a buscar modos alternativos de resoluções de litígios (conciliação, mediação ou arbitragem).⁵⁰

98. A abordagem que já existe em muitos sistemas judiciais de harmonização de decisões em muitas áreas (divórcio, demissão, lesão) poderia ser revitalizada por meio de um procedimento probabilístico ou atuarial.⁵¹ Foram criados serviços alternativos de resolução de litígios on-line para auxiliar na avaliação do montante de compensação para pequenos litígios. No entanto, estas abordagens interessantes não são imparciais e não devem privar cidadãos do acesso a um juiz ou contestar o princípio do contraditório.

Experimentos realizados na França

Por iniciativa do Ministério da Justiça, os dois tribunais de recurso em Rennes e Douai concordaram em testar uma versão de software, apresentado como “previsibilidade da justiça”, em vários recursos de litígio, na primavera de 2017, que na realidade foi uma análise de decisões civis, sociais e comerciais de todos os tribunais franceses de apelação.

Embora estes dados internos e exaustivos da jurisprudência já estivessem disponíveis para eles gratuitamente por muitos anos (banco de dados JURICA), o Ministério os fez especialmente disponíveis para a editora quando se ofereceram para realizar uma análise quantificada (inovadora) das somas atribuídas pelos dois tribunais, bem como uma classificação geográfica das discrepâncias encontradas em julgamentos semelhantes.

O objetivo declarado do software era, portanto, criar uma ferramenta de apoio para reduzir, se necessário, a excessiva variabilidade nas decisões judiciais, em nome do princípio da igualdade dos cidadãos perante a lei.

O resultado do experimento foi contraditoriamente debatido entre os dois tribunais de recurso. O Ministério da Justiça e a empresa de tecnologia jurídica que projetou o produto, infelizmente, declararam a ausência de valor agregado à versão testada do software para o trabalho de reflexão e tomada de decisão

⁵⁰ A esse respeito, ver o relatório de informações nº 495 (2016-2017) preparado em nome da comissão de leis do Senado, e apresentada em 4 de abril de 2017, pelo senador Philippe Bas.

⁵¹ Com relação à abordagem atuarial, a oferta da Case Law Analytics anuncia a avaliação de risco de uma solução numa disputa.

dos magistrados.

Mais significativamente, foram revelados vieses de raciocínio de software que levaram a aberrações ou resultados inadequados devido à confusão entre simples ocorrências lexicais de raciocínio judicial e as causalidades que tinham sido decisivas no raciocínio dos juízes.

6.1. Uma nova ferramenta de computação para calcular escalas

99. Os procedimentos para calcular escalas em várias matérias civis (por exemplo, indenização por lesão corporal, indenização compensatória e rescisão) parecem ser consideravelmente melhorados quando combinados com outras técnicas de processamento, sujeitas a numerosas medidas de projeto e uso (efeito performativo).⁵²

100. É importante destacar o que Jean-Paul Jean, presidente do grupo de trabalho de avaliação do O CEPEJ, chamou de “desafio qualitativo” em um simpósio sobre dados abertos realizado em 2016, na França: o procedimento por aprendizado de máquina ou qualquer outro método de processamento deve usar certificados originais, cuja integridade foi verificada e que foram enriquecidos para distinguir o importante do insignificante.⁵³

101. O outro risco já encontrado pelos projetistas dessas ferramentas é o de “espionagem de dados”, ou seja, a seleção de dados somente de fluxo ascendente que sejam significativos para grids de análise predeterminados, por exemplo, excluindo-se amostras de decisões que se prestam mal a correlações de sequências linguísticas através de aprendizado de máquina ou qualquer outro método (por exemplo, decisões sem apresentação do litígio ou pouco motivadas).

102. Mas se calcularmos uma escala, já não estamos fazendo uma pequena previsão? A fronteira pode parecer relativamente tênue, a menos que se distinga claramente o objetivo do processo: o objetivo aqui não é fornecer informações prescritivas, mas fornecer informações sobre um estado de coisas.

⁵² O efeito performativo ou de auto realização é o risco de que um sistema produza o mesmo resultado progressivamente, influenciando os produtores de informação de entrada; este efeito é frequentemente mencionado em relação às escalas judiciais que, quando informadas por decisões baseadas nessas escalas, tendem a ser representativas apenas de si mesmas.

⁵³ J-P. Jean, “Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence”, Simpósio de 14 de outubro de 2016 realizada no Tribunal de Cassação, <https://www.courdecassation.fr/IMG///Open%20data,%20par%20Jean-Paul%20Jean.pdf>, página consultada em 17 de março de 2018

103. Sujeito a estas precauções metodológicas e operacionais, as escalas estão disponíveis para lidar com certas disputas e são ferramentas poderosas para harmonização da jurisprudência. As escalas foram previamente calculadas com base em amostras mais ou menos limitadas de decisões, mas as ferramentas de IA possibilitam pesquisar mais decisões e podem produzir escalas mais precisas, aplicando um padrão e, portanto, dando mais peso aos resultados.

6.2. Resolução de disputa on-line

104. Todos os tribunais europeus lidam, em diferentes graus, com processos civis repetitivos de baixo valor. A ideia de facilitar o procedimento através da tecnologia da informação e/ou terceirizá-los dos tribunais é amplamente compartilhada. Grã-Bretanha, Holanda e Letônia são exemplos de países que já têm implementado ou estão prestes a implementar esses tipos de soluções mais ou menos automatizadas.⁵⁴ Para os pedidos transfronteiriços, a União Europeia fixou, através do Regulamento n.º 524/2013, um quadro comum disponível na Internet (“pequenas ações europeias”).

105. No entanto, o âmbito destes serviços de resolução de litígios online (online dispute resolution- ODR) parece ter evoluído progressivamente. Eles passaram de serviços on-line limitados para medidas alternativas de resolução de litígios, antes da queixa ser apresentada perante o tribunal, e agora estão sendo introduzidos cada vez mais no próprio processo judicial, a ponto de oferecer serviços judiciais.⁵⁵ Eles não tratam apenas de disputas de baixo valor, mas também se aplicam a litígios fiscais ou relativos aos serviços de seguridade social e processos de divórcio.

106. Para aqueles que defendem tais soluções, que são de interesse para várias profissões jurídicas e para o setor privado, o acesso à justiça poderia ser significativamente melhorado por meio de uma ampla solução combinando ODR e IA (ou pelo menos sistemas especialistas, veja o capítulo 3 acima para a

⁵⁴ eja, por exemplo, o sistema de resolução de disputas on-line na Grã-Bretanha - <https://www.judiciary.gov.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version1.pdf>. Veja também o sistema PAs na Holanda, que emite decisões automatizadas com base em autorizações anteriormente concedidas e que deram origem a litígios a nível nacional e perante o EJC: há dois casos (c-293/17 e c-294/17) levados ao Conselho de Estado dos Países Baixos (agricultores / conservação da natureza contra os Países Baixos) para determinar se um sistema (Programa Regulando Nitrogênio) pode ou não decidir se os fazendeiros estão violando a lei. Recentemente o Tribunal de Justiça da União Europeia em Luxemburgo respondeu aos pedidos de decisão prejudicial relativa a estes processos (ECLI: EU: C: 2018: 882)

⁵⁵ Darin Thompson, “Creating new pathways to justice using simple artificial intelligence and online dispute resolution”, Osgoode Hall Law School of York University.

distinção). A ideia é acompanhar os reclamantes em um diagnóstico automatizado dos litígios através de uma série de questões tratadas pela máquina, resultando em propostas de solução. O trabalho do laboratório Cyberjustice de Montréal, que reúne as várias fases de pré-contencioso e contencioso em um processo informatizado baseado em disputas de baixa intensidade (por exemplo, pequenas causas judiciais dos tribunais de Quebec), é uma boa ilustração da hibridização.⁵⁶ De acordo com os desenvolvedores, há benefícios claros em termos de eficiência e qualidade.

107. Mas em que base seria calculada a possível compensação proposta por tal sistema? Qual método? O algoritmo trata a informação de forma justa? As propostas têm vocação para poder ser discutidas contraditoriamente com a ajuda de um terceiro formado e certificado? O acesso ao juiz é sempre possível? Alguns autores chegam a ver o uso disseminado desses métodos de resolução de disputas como uma nova manifestação do “solucionismo” digital, ou seja, o uso sistemático de tecnologias para tentar resolver problemas não é necessariamente adequado.⁵⁷ Deve-se notar também que, na Europa, um quadro regulamentar vinculativo para os Estados-Membros foi recentemente implementado: o Artigo 22 do Plano de Monitoramento de Segurança de Dados explicitamente prevê que as pessoas possam recusar-se a ser objeto de uma decisão baseada exclusivamente no processamento automatizado, com certas exceções.⁵⁸

56 <http://www.cyberjustice.ca/projets/odr-plateforme-daide-au-reglement-en-ligne-de-litiges/>

57 Evgeny Morozov, “Pour tout résoudre cliquez ici”, FYP editions, cited by David Larrousserie, “Contre le ‘solutionnisme’ numérique”, *Le Monde*, 6 de outubro de 2014, https://www.lemonde.fr/sciences/article/2014/10/06/contre-le-solutionnisme-numerique_4501225_1650684.html

58 Artigo 22º, nº 1, do Regulamento da UE 2016/679: “O titular dos dados tem o direito de não se sujeitar a uma decisão baseada exclusivamente no processamento automatizado”; estão previstas exceções (como consentimento da pessoa em litígio), mas “medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses da pessoa em litígio “devem ser implementados pelo controlador do processo, “incluindo o direito da pessoa em litígio de obter intervenção humana da pessoa responsável pelo tratamento, expressar o ponto de vista e contestar a decisão”. Ver, no mesmo sentido, a Convenção do Conselho da Europa para a Proteção dos Indivíduos relativamente ao tratamento automático de dados pessoais, com a redação que lhe foi dada pelo Protocolo de maio de 2018, quando este entrar em vigor. O §1º, alínea a, do artigo 9º, prevê o princípio de que “toda a pessoa tem o direito de não estar sujeita a uma decisão que a afete significativamente, que seria tomada com base apenas no processamento automatizado de dados, sem que seu ponto de vista tenha sido levado em conta”. Não obstante este princípio de proibição, o §2º do artigo 9º afirma que “o parágrafo 1º não se aplicará se a decisão for autorizada por uma lei a que o controlador está sujeito e que também prevê medidas apropriadas para salvaguardar os direitos, liberdades e interesses legítimos da pessoa em causa”.

O relatório explicativo afirma (§75): “É essencial que qualquer pessoa que esteja sujeita a uma decisão puramente automatizada deva ter o direito de contestar essa decisão apresentando seu ponto de vista e seus argumentos. Em particular, o sujeito dos dados deve ter a possibilidade de provar a possível inexactidão dos dados pessoais antes de seu uso, a inadequação do perfil a ser aplicado a sua situação particular ou outros fatores que terão impacto no resultado da decisão automatizada. Isto é particularmente o caso em que a aplicação de um raciocínio algorítmico, levando à limitação de um direito, a recusa de um benefício social ou a avaliação de sua capacidade de contrair empréstimo apenas como em software, tem o efeito de estigmatizar indivíduos. No entanto, o sujeito

108. Os benefícios potenciais de um sistema de ODR devem, portanto, ser medidos caso a caso. Seu grau de integração em um processo jurisdicional completo (do pré-litígio ao contencioso real) bem como o papel mais ou menos decisivo da IA na execução do processo.

109. O ODR já oferece conhecimento a montante dos processos judiciais. Seu papel é claramente contribuir para a implementação da conciliação, mediação e serviços de arbitragem fora do tribunal. Esses serviços também podem ser usados durante o processo contencioso sob a supervisão de juizes antes que eles decidam sobre o resultado dos litígios com base no mérito (Essa fase é por vezes tornada obrigatória para certos litígios).

110. Por outro lado, a contribuição real da IA deve ser avaliada. É apenas uma questão de usar o aprendizado de máquina para estabelecer escalas indicativas ou prescrever uma solução? É realmente a IA que está sendo usada ou um sistema especialista ou apenas uma sequencia simples de regras lógicas? Em qualquer caso, deve ser possível combinar esses sistemas com os requisitos de transparência, neutralidade e lealdade.⁵⁹

111. Por último, é igualmente necessário examinar a forma como os demandantes são encorajados a usar o sistema: existe uma confusão potencial no próprio nome do que está em oferta? Se alguém fala de um tribunal, deve ser a forma de organização definida pela Convenção Europeia dos Direitos Humanos e não simplesmente uma instituição de justiça privada com a mera aparência justiça estatal.⁶⁰ O recurso a um juiz é claramente possível? Na Holanda, os contratos de seguro de saúde privado Parecem, por padrão, fornecer um ODR antes de qualquer ação legal ser apresentada.

não pode exercer este direito se a decisão automatizada estiver prevista na lei, que prevê medidas apropriadas para salvaguardar os direitos, liberdades e interesses legítimos da pessoa em causa”. Ver também o artigo 9º, nº 1, alínea c, da Convenção modernizada 108, que prevê o direito do titular dos dados “obter, a seu pedido, conhecimento do raciocínio subjacente à processamento de dados, quando os resultados de tal processamento são aplicados a ele”. O relatório explicativo da Convenção modernizada (§77) afirma: “Os titulares de dados têm o direito de obter conhecimento do raciocínio subjacente ao processamento de dados, incluindo as consequências deste raciocínio e as conclusões que podem ser extraídas, especialmente quando se utilizam algoritmos para tomadas de decisão automatizadas, em particular como parte do perfil. Por exemplo, no caso de um sistema de classificação de crédito, os mutuários têm o direito de conhecer a lógica por trás do processamento de seus dados que leva à decisão de conceder ou recusar crédito, em vez de simplesmente ser informado da decisão em si. A compreensão desses elementos contribui para o exercício efetivo de outras salvaguardas essenciais, como o direito de oposição e o direito de recurso à autoridade competente”.

59 Charlotte Pavillon, “Preocupações com um juiz digital”, nrc.nl, <https://www.nrc.nl/nieuws/2018/01/19/zorgen-om-populaire-digitale-rechter-a1588963>

60 Scarlett-May Ferrié, Algoritmos testados contra um julgamento justo, documento consultado em <http://lexis360.fr>, downloaded on 09/07/2018, § 27-38

112. O grupo de trabalho de mediação do CEPEJ (CEPEJ-GT-MED) iniciou em 2018 reflexões sobre as contribuições das tecnologias da informação para a resolução de litígios com métodos alternativos. O CDCJ está atualmente conduzindo um trabalho profundo sobre os ODRs, a fim de identificar as potencialidades dessas ferramentas, mas também os pontos problemáticos que possam resultar em possíveis violações dos artigos 6º, 8º e 13º da Convenção Europeia dos Direitos Humanos.

6.3. As principais garantias a serem reafirmadas no contexto de processos civis, comerciais e administrativos

Direito de acesso a um tribunal

113. O fornecimento de ferramentas de resolução de litígios online não deve afetar o direito de acesso a um tribunal na acepção do artigo 6º, mesmo que esse direito não seja absoluto e possua limitações implícitas.⁶¹ Em matéria civil, por exemplo, todo litigante tem o direito de submeter a um tribunal qualquer litígio relacionado aos seus “direitos e obrigações civis”.⁶² Em 2015, a Assembleia Parlamentar do Conselho da Europa aprovou uma resolução sobre “Acesso a justiça e a Internet: potencialidades e desafios” nos quais convocou para garantir que “as partes envolvidas em procedimentos de ODR mantenham o direito de acesso a um recurso que satisfaça os requisitos de um julgamento justo de acordo com o Artigo 6º da Convenção⁶³”.

Princípio do contraditório

114. No que diz respeito às escalas, parece imperativo torná-las acessíveis aos cidadãos e, repassando certa quantidade de informação quantitativa (número de decisões processados para obter a escala, por exemplo) e qualitativa (origem das decisões, representatividade das amostras selecionadas, distribuição de decisões entre diferentes critérios, tais como contexto econômico e social), para que possam compreender o significado das escalas construídas, medir

61 Art.6 §1 1. Todos têm direito a uma audiência justa e pública. Por uma autoridade independente e um tribunal imparcial... quem decidirá (...) o mérito de qualquer acusação criminal contra ele. para as limitações, veja Deweer c. Belgique, § 49; Kart c. Turquie [GC], § 67.

62 CEDH, Golder c. Royaume-Uni, §§ 28-36

63 Resolução 2054 (2015) da Assembleia Parlamentar do Conselho da Europa (APCE), 10 de novembro de 2015, <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-EN.asp?fileid=22245&lang=en>

possíveis limites e poder debatê-los perante um juiz.

Igualdade de armas

115. A utilização de meios tecnológicos não deve causar desequilíbrios entre as partes, uma vez que a utilização de meios digitais pode, de fato, facilitar os procedimentos para certos operadores (instituições, empresas com recursos, pessoas com conhecimentos de informática) e, ao contrário, colocar dificuldades a certos indivíduos que são menos familiarizados com computadores. É importante que nenhum indivíduo seja deixado sozinho na frente de suas telas, e que eles sejam informados de que podem procurar aconselhamento jurídico e são assistidos sempre que necessário.

Imparcialidade e independência dos juízes

116. É questionável se a norma derivada da tendência majoritária mencionada anteriormente na seção 2.2. não produziria efeitos indiretos sobre a independência e imparcialidade dos juízes, e particularmente em sistemas onde a independência do judiciário não é plenamente alcançada. Nestes sistemas, existe o risco de que este 'padrão numérico' possa ter um efeito indireto de pressão sobre os juízes durante a tomada de decisão ou que o executivo monitore aqueles que derrogam a norma.

Direito de aconselhar

No início deste capítulo, mencionamos as vantagens derivadas da aplicação de ferramentas de justiça preditiva para advogados e, em particular, a possibilidade de fornecer aos seus clientes um melhor aconselhamento informativo, avaliando empiricamente e sistematicamente as chances de sucesso de um procedimento. No entanto, vamos imaginar um caso em que as chances de sucesso para o litigante são extremamente pobres: isso poderia afetar a decisão do advogado de ajudar seu cliente? A prática profissional deve ter como objetivo minimizar o risco de que as pessoas que buscam aconselhamento jurídico possam, em última instância, ser privado dele.

7. Questões específicas da justiça criminal: prevenção de infrações, risco de reincidência e avaliação de periculosidade

Mesmo eles não sendo especificamente projetados para serem discriminatórios, o uso de estatísticas e IA em processos criminais tem mostrado um risco de provocar o ressurgimento de doutrinas deterministas em detrimento das doutrinas de individualização das sanções, que foram amplamente adquiridas desde 1945 na maior parte dos sistemas judiciários europeus.

117. A utilização de ciências e tecnologias da IA em matéria penal coloca desafios específicos, porque a sua aplicação poderia refletir alguns debates públicos sobre o alegado comportamento ofensivo previsível. Este debate, no entanto, parecia ter sido completamente resolvido durante trinta anos em vários países europeus. Na Itália, por exemplo, o artigo 220, § 2º, do Código de Processo Penal, exclui expressamente o uso de opinião de especialista para estabelecer “o hábito ou profissionalismo no crime, a tendência a cometer um crime, o caráter e personalidade do acusado e, em geral, as qualidades do acusado, independentemente de causas patológicas”. Na França, a doutrina chamada “nova defesa social”, desenvolvida por Marc Ancel, foi o fundamento do direito penal: em vez de uma abordagem meramente punitiva e determinista, foi introduzido um sistema de reabilitação social para impedir a prática de um delito, evitando as condições para o crime. Essa abordagem é compartilhada por vários instrumentos de políticas criminais que se concentram nos objetivos de reeducar e reintegrar ofensores.⁶⁴

118. Por conseguinte, os instrumentos de justiça penal devem ser concebidos de acordo com os princípios fundamentais da reabilitação,⁶⁵ incluindo o papel do juiz na individualização da sentença, com base em elementos objetivos da personalidade (formação, emprego, apoio médico-social), sem qualquer outra forma de análise do que aquela realizada por técnicos especificamente treinados, como oficiais da condicional, por exemplo. Técnicas de análise de big data poderiam ser utilizadas por esses profissionais para centralizar e coletar informações sobre a pessoa acusada de um crime ou contravenção, que poderia então ser encontrada em várias instituições e agências. Informações que precisariam ser examinadas por um juiz, às vezes dentro de um prazo muito curto (por exemplo, no contexto de procedimentos acelerados para acórdão).

⁶⁴ veja European Court of Human Rights, Grand Chamber, Vinter and Others vs. United Kingdom, paras. 114-118

⁶⁵ Por outro lado, o uso de IA para fins de tratamento e reabilitação (por exemplo, para coletar dados do tratamento administrado ou métodos de reintegração na prisão) deve ser encorajado.

7.1. As ferramentas utilizadas pelo serviço de investigação antes do julgamento criminal

119. Os instrumentos descritos como “policiamento preditivo” (antes do processo judicial ou antes de um encaminhamento judicial) já estão crescendo rapidamente e estão começando a ser conhecidos pelo público em geral (pense, por exemplo, nas “listas de não voo”, que são na verdade uma aplicação de análise “big data” que coleta e analisa dados sobre potenciais terroristas com o propósito de impedir a prática de atos ou os algoritmos usados para detectar fraudes ou lavagem de dinheiro).

120. Em geral, um grande número de ferramentas de computador é comumente usado para impedir a prática de atos criminosos (identificando possíveis locais onde isso pode acontecer ou seus autores) ou persegui-los com maior eficácia.⁶⁶ A primeira categoria inclui instrumentos de “policiamento preditivo” que são usados para prevenir certos tipos de infrações com regularidade em sua ocorrência, como roubo, violência nas ruas, roubo de veículos. A designação dessas ferramentas deriva de sua capacidade de determinar precisamente onde e quando essas ofensas poderiam ser cometidas e reproduzir essas informações em um mapa geográfico na forma de “hot spots” que são monitorados em tempo real por patrulhas policiais. Esse processo é chamado de mapeamento criminal preditivo. A maioria dos softwares usados nesta área é baseada em evidências históricas de localização do crime, tais como relatórios policiais, mas novas tecnologias ainda mais poderosas combinam dados variados e diferentes fontes também estão sendo testadas.⁶⁷ Esses instrumentos, que exibem taxas de eficácia muito persuasivas, também possuem efeitos dissuasivos sobre a prática de infrações nas áreas ao redor dos “hot spots”, gerando um efeito positivo para as políticas públicas.⁶⁸

⁶⁶ Ver Ales Zavrnsnik, Big Data, crime and social control, página 194 e seguintes, que lista em detalhes uma série de instrumentos utilizados pelos serviços policiais na Europa e nos Estados Unidos.

⁶⁷ Por exemplo, como parte do projeto “E-Security - ICT para segurança urbana baseada em conhecimento e preditiva” (<http://www.esecurity.trento.it/>), que foi realizado na cidade de Trento, na Itália, de novembro de 2012 a maio de 2015, um banco de dados contendo informações sobre os crimes denunciados à polícia, os resultados dos inquéritos realizados pela câmara municipal sobre vitimização e segurança real e percebida pelos cidadãos, elementos da desordem física e social urbana da polícia, bem como outras variáveis relacionadas ao “SmartCity” (por exemplo, informações sobre condições sócio demográficas, ambiente urbano, iluminação noturna, a presença de câmeras de vigilância, transporte público etc.) “ Foi criada para melhor equipar o trabalho de prevenção ao crime e melhoria da segurança urbana. Os funcionários do projeto comprovaram a confiabilidade das técnicas utilizadas, que se diziam possibilitar a previsão de atos criminosos com um taxa de sucesso de cerca de 60-65% e que ajudaria a combater melhor a criminalidade num contexto de escassez geral de recursos. Além disso, testes conduzidos no Reino Unido como parte de um projeto piloto destinado a prever os locais de possíveis arrombamentos, roubos e assaltos mostram que as projeções do software utilizado, denominado PREDPOL, foram precisos em 78% dos casos, em comparação com 51% das técnicas tradicionais.

⁶⁸ A indicação de uma concentração geográfica do crime ajudaria as forças policiais a questionar melhor seus fatores

121. No entanto, os recursos preditivos dessas ferramentas demonstram suas limitações relativas a crimes de natureza menos regular ou que visem diferentes locais, como o terrorismo. Além disso, uma das suas fraquezas é o efeito de “círculo vicioso” e “profecias autorrealizáveis”: bairros considerados em risco atraem mais atenção da polícia e a polícia detecta mais crimes, o que leva a vigilância policial excessiva dos habitantes dessas comunidades.⁶⁹ Por fim, perguntas sobre uma possível “tirania do algoritmo” que poderia minimizar ou mesmo substituir progressivamente o julgamento humano não está totalmente ausente dentro dos próprios serviços policiais, ainda que, por enquanto, a tecnologia seja apresentada como permanecendo ao serviço da polícia para melhor equipá-los na tomada de decisões.⁷⁰

122. Além disso, a análise de big data está sendo cada vez mais aplicada no processo criminal. Ferramentas como o Connect, que é usada pela polícia do Reino Unido para analisar bilhões de dados gerados em transações financeiras para encontrar correlações ou padrões de operações, ou o Banco de Dados Internacional de Exploração Sexual de Crianças (ICSE DB), gerenciado pela Interpol, que ajuda a identificar as vítimas e /ou autores através da análise, por exemplo, de móveis e outros objetos presentes em imagens abusivas, ou a análise de ruídos de fundo em vídeos, provaram ser particularmente eficazes no combate ao crime. Com o Connect, por exemplo, as pesquisas que anteriormente exigiam meses de investigação agora podem ser realizadas em minutos, com um alto nível de complexidade e volume de dados.

123. A doutrina, no entanto, questiona a lógica gerencial da resposta ao crime fornecida por essas ferramentas preditivas, em que uma análise das razões para o crime se torna menos importante do que fazer algo aqui e agora. Isso está ocorrendo em um momento em que os orçamentos disponíveis estão encolhendo e a polícia deve fornecer o mesmo nível de proteção pública, mas com pessoal, equipamento e recursos limitados.⁷¹

ambientais predisponentes na área identificada (por exemplo, iluminação, presença de lojas, etc.) e fornecer respostas adequadas ao serem consultadas.

69 “Predicting crime, LAPD style”, The Guardian, 25 de junho de 2014.

70 “How technology is allowing police to predict where and when crime will happen”, The Independent, 7 de outubro de 2017

71 Ales Završnik, Big Data, crime and social control, página 196

7.2. As ferramentas durante o julgamento criminal

124. O uso de ferramentas preditivas⁷² por juízes em julgamentos criminais é muito raro na Europa.

125. O HART (Harm Assessment Risk Tool) foi desenvolvido em parceria com a Universidade de Cambridge e agora está sendo testado no Reino Unido. Esta tecnologia com base no aprendizado de máquina foi testada usando arquivos da polícia de Durham, entre 2008 a 2012. Aprendendo com as decisões tomadas pelos policiais durante este período, e se certos suspeitos são ou não reincididos, espera-se que a máquina seja capaz de avaliar o risco - baixo, médio ou alto - de suspeitos, com base em cerca de trinta fatores, alguns dos quais não estão relacionados ao crime cometido (por exemplo, código postal e sexo).

126. Em testes realizados inicialmente em 2013, durante os quais o comportamento do suspeito foi observado durante um período de dois anos após o cometimento do crime, as previsões do HART foram desenvolvidas para ser 98% eficaz na previsão de baixo risco e 88% eficaz em alto risco de reincidência. Nesta fase experimental, o HART terá um valor puramente consultivo para o juiz. Além disso, as auditorias do funcionamento do HART e a confiabilidade de suas conclusões serão conduzidas regularmente pela polícia.

127. Embora esta seja a única ferramenta preditiva identificada na Europa até o momento, ela oferece a oportunidade de questionar os desafios que os julgadores poderiam enfrentar em um futuro próximo, se esse tipo de aplicação for testado em um ambiente de maior escala, particularmente baseada na experiência americana.

128. Nos Estados Unidos⁷³, a ONG ProPublica revelou os efeitos discriminatórios do algoritmo usado em COMPAS (Perfil de Gerenciamento Corretivo de Infratores para Sanções Alternativas), a ambição é avaliar o risco de reincidência quando o juiz determinar a sentença de um indivíduo.

129. Este algoritmo, que foi desenvolvido por uma empresa privada e que deve ser usado por juízes em certos estados federais americanos, inclui 137 perguntas respondidas pelo réu ou informações retiradas de registros criminais.

⁷² Na literatura, essas ferramentas são freqüentemente chamadas de “justiça algorítmica” ou “justiça automática”ou “justiça simulada”.

⁷³ Um estudo de 2015 identificou cerca de sessenta ferramentas preditivas nos Estados Unidos

As perguntas são bastante variadas e incluem a presença de um telefone em casa, dificuldade em pagar contas, histórico familiar, antecedentes criminais do acusado, etc.⁷⁴ O algoritmo classifica a pessoa em uma escala de 1 (baixo risco) a 10 (alto risco). Isso representa uma ajuda para a tomada de decisões judiciais, sendo suas conclusões apenas uma das variáveis consideradas pelo juiz ao decidir sobre a sentença.

130. Às populações afro-americanas foi atribuída uma taxa de alto risco de reincidência, o dobro de outras populações no prazo de dois anos após a condenação. Este efeito não foi naturalmente procurado pelos desenvolvedores.⁷⁵ Em contraste, o algoritmo considerou que outras populações pareciam muito menos propensas a repetir um crime.⁷⁶ É claro que deve ser notado que este tipo de interpretação enganosa, na realidade, revela apenas a fragilidade social e econômica de certos grupos de populações que obviamente não são criminogênicas por natureza. Pesquisadores do Dartmouth College também mostraram que esse tipo de algoritmo não produz valor acrescentado, uma vez que pessoas sem antecedentes criminais podem reproduzir exatamente a mesma avaliação simplesmente respondendo ao questionário.

131. Além disso, a falta de transparência nos processos de operação de algoritmos projetados por empresas privadas (que reivindicam propriedade intelectual) foi outro motivo de preocupação. Se levamos em conta o fato de que eles utilizam os dados de origem das próprias autoridades estaduais, sua

74 Existem outros algoritmos que foram desenvolvidos com base em observações críticas expressas pela doutrina (veja o próximo capítulo) que são baseados em variáveis mais restritas, mais diretamente relacionadas ao crime cometido e menos relacionada à raça, gênero ou status socioeconômico. Um exemplo é a Ferramenta de Avaliação de Segurança Pública, utilizada em 30 jurisdições americanas.

75 Este efeito puramente discriminatório pode, de fato, ser explicado pela “calibração” relativamente permissiva do modelo de algoritmo, que cria muitos “falsos positivos”.

76 As populações negras eram mais frequentemente classificadas como de alto risco, enquanto não reincidiam dentro de dois anos da condenação; populações brancas, por outro lado, que eram mais frequentemente classificadas como de baixo risco, voltavam a praticar dentro de dois anos. Em resumo, o algoritmo superestimou o risco de reincidência para negros e subestimou para brancos (os “falsos positivos” eram na maioria negros, enquanto os “falsos negativos” eram brancos). Em resposta às alegações da ProPublica, a NorthPointe (agora Equivant, desde essa controvérsia) respondeu que as populações brancas e negras eram igualmente representadas considerando os “verdadeiros positivos”, ou seja, aqueles que realmente reincidiram. A questão de como conciliar tanto a precisão do algoritmo na detecção de reincidência como a necessidade de evitar efeitos discriminatórios para populações negras tem sido fonte de intenso debate na literatura; ver em particular Chouldechova A (2016), “Uma previsão justa com um impacto dispar: um estudo sobre viés em instrumentos de previsão de reincidência”, disponível em <http://arxiv.org/abs/1610.07524>; e também “O preconceito em escores de riscos criminais é matematicamente inevitável, dizem os pesquisadores”, disponível em <https://www.propublica.org/article/bias-in-riminal-risk-scores-is-mathematically-resistible-researchers-ay>. Este debate também reflete a questão da legitimidade de uma empresa privada, sem qualquer controle institucional, para arbitrar entre dois requisitos opostos: o de defender a sociedade, por um lado, e de respeitar os direitos dos indivíduos, por outro.

falta de prestação de contas aos cidadãos representa um grande problema democrático. O público é informado sobre operações de big data acidentalmente, esporadicamente e quando há vazamentos ou erros: um exemplo disso foi quando a ProPublica revelou as falhas no algoritmo COMPAS. Diante da recusa da empresa proprietária, a ONG teve que apelar às autoridades públicas para acessar os dados e contratar seu próprio cientista para examinar o algoritmo.

7.3. Os desafios da “previsão” em matéria penal

132. Vimos nas seções anteriores que o grau de desenvolvimento de ferramentas preditivas na Europa no campo criminal é muito variado. Apesar dos instrumentos descritos como “policiamento preditivo” estarem crescendo rapidamente e começando a chamar a atenção do público em geral, a situação não é a mesma quando se trata da aplicação desse tipo de ferramenta por juízes em julgamentos criminais. No que diz respeito aos instrumentos disponíveis para serviços judiciais, considerações sobre suas vantagens e desvantagens já foram expressas anteriormente. Vamos agora estudar as ferramentas específicas para julgamentos criminais.

133. Em primeiro lugar, é importante excluir argumentos baseados apenas na eficiência ou ineficiência dessas ferramentas. Os exemplos ilustrados acima mostram que pode haver grandes oportunidades, mas também riscos reais na aplicação de novas tecnologias, por falta de precauções necessárias. Os julgadores e os atores judiciais devem ser particularmente vigilantes e desempenhar papel ativo no desenvolvimento dessas tecnologias; é necessário um monitoramento contínuo para determinar a sua real eficácia e eficiência, e evitar consequências imprevistas. Isso é ainda mais importante em processos criminais por causa de seu impacto direto na liberdade pessoal dos indivíduos.⁷⁷

⁷⁷ Um extrato da decisão da Suprema Corte de Wisconsin, caso Wisconsin vs. Loomis, pode oferecer inspiração também a nível europeu: “É importante considerar que ferramentas como o COMPAS continue a mudar e evoluir. As preocupações que abordamos hoje podem muito bem ser aliviadas no futuro. Cabe ao sistema de justiça criminal reconhecer que, nos próximos meses e anos, dados adicionais de pesquisa estarão disponíveis. Diferentes e melhores ferramentas podem ser desenvolvidas. À medida que os dados mudam, nosso uso de ferramentas baseadas em evidências tem que mudar também. O sistema de justiça deve acompanhar a pesquisa e continuamente avaliar o uso dessas ferramentas.”

134. Isto implica que tanto os benefícios como os inconvenientes da aplicação de tais ferramentas no campo judicial devem ser cuidadosamente avaliados.

135. Os defensores frequentemente argumentam que elas são neutras e que confiariam métodos factuais e objetivos que ajudam a tornar a justiça mais precisa e transparente. Outro grande trunfo, afirma-se, é a sua eficiência, que às vezes excede as capacidades humanas, sendo extremamente valiosa em um contexto geral de redução de fundos públicos ou mesmo de escassez de recursos.

136. Por outro lado, os seus detratores expressam preocupação sobre o impacto e perigo potencial que essas ferramentas teriam numa justiça individualizada e igual para todos: a inclusão de variáveis algorítmicas como o histórico criminal e histórico familiar significa que o comportamento passado de certo grupo pode decidir o destino de um indivíduo, que é, naturalmente, um ser humano único com um contexto social específico, educação, habilidades, grau de culpa e possui motivações específicas que o levou a cometer crime.⁷⁸ Eles também argumentam que as decisões humanas podem basear-se em valores e considerações (por exemplo, sociais) que não seriam retidas pela máquina. Por exemplo, um juiz pode decidir ordenar a fiança de uma mulher infratora que tem um risco de reincidência, com base numa hierarquia de valores, definindo maior valor pelo seu papel de mãe e protetora de seus filhos, enquanto o algoritmo seria capaz de determinar o risco de reincidir com mais precisão, mas não seria capaz de operar tal hierarquia de prioridades.

137. Em matéria penal, existem também riscos potenciais de discriminação quando considera-se que essas ferramentas, que são construídas e interpretadas por humanos, podem reproduzir desigualdades injustificadas e já existentes no sistema de justiça criminal em causa; em vez de corrigir certas políticas problemáticas, a tecnologia pode acabar legitimando-as. Como já mencionado, a ONG ProPublica⁷⁹ revelou claramente os efeitos discriminatórios do algoritmo usado no COMPAS⁸⁰, que previa que as populações negras eram duas vezes mais propensas a reincidir, no comparativo com as populações brancas, nos dois anos seguintes à aplicação da sentença. Eles poderiam, no entanto, ajudar a revelar

78 Aleš Završnik, "Big Data, crime and social control", pagina 196

79 www.propublica.org/article/technical-response-to-northpointe.

80 Outros algoritmos concentram-se em outros elementos mais diretamente relacionados à infração cometida.

os erros na tomada de decisão e indicar que eles podem ser corrigidos.⁸¹ Além disso, a falta de transparência do algoritmo no processo de construção por suas empresas proprietárias e a ausência da prestação de contas com o público é motivo de preocupação.

138. À luz do exposto, quando os algoritmos são usados no contexto de um julgamento criminal, parece essencial garantir plenamente o respeito ao princípio da igualdade de armas e presunção de inocência, estabelecidos pelo artigo 6º da CEDH. A parte interessada deve ter acesso e poder contestar a validade científica de um algoritmo, a ponderação dada a seus vários elementos e quaisquer conclusões errôneas a que se refere, sempre que um juiz sugere que ele / ela possa usá-lo antes de tomar sua decisão. Além disso, esse direito ao acesso é também abrangido pelo princípio fundamental da proteção de dados pessoais. Todas as pessoas têm o direito de não estarem sujeitas a decisões que as afetam significativamente apenas com base no processamento automatizado de dados, sem que o seu ponto de vista seja levado em consideração.

139. A este respeito, existe uma diferença entre a Europa e os Estados Unidos no que diz respeito ao direito de acesso ao algoritmo: se nos Estados Unidos as autoridades judiciais ainda relutam em reconhecer plenamente esse direito e encontrar um equilíbrio entre interesses privados (particularmente a proteção da propriedade intelectual) e os direitos da defesa, na Europa o quadro é mais protetor por causa do GDPR, que estabelece o direito à informação sobre a lógica subjacente das decisões tomadas usando algoritmos.⁸²

140. As considerações expressas anteriormente sobre os efeitos potencialmente negativos dessas ferramentas na imparcialidade do juiz também são válidos em matéria penal: um juiz que decide contra a previsão de um algoritmo é provável assumir riscos, pois ele assume maior responsabilidade. Não parece irrealista imaginar que os juízes relutariam em assumir essa carga adicional, particularmente em sistemas onde seus mandatos não são permanentes, mas sujeitos a voto popular,⁸³ ou em que podem responder pessoalmente (disciplinar, civil ou mesmo criminal), especialmente se as suas garantias estatutárias em matéria disciplinar forem insuficientes.

81 Mojca M. Plesnicar and Katja Sugman Stubbs, "Subjectivity, algorithms and the courtroom".

82 Artigo 15º, nº 1, alínea h), do Regulamento UE 2016/679: "O titular dos dados tem o direito de obter do controlador "... as seguintes informações: "... a existência de tomada de decisão, incluindo a definição de perfis, como referido nos nºs 1 e 4 do artigo 22º, e pelo menos nesses casos, informações significativas sobre a lógica envolvida, bem como a importância e as consequências previstas de tal processamento para a pessoa em causa".

83 Mojca M. Plesnicar and Katja Sugman Stubbs, "Subjectivity, algorithms and the courtroom"

8. Questões específicas relacionadas com a proteção de dados pessoais

O uso de algoritmos levanta a questão da proteção de dados pessoais quando processados. O princípio de precaução deve ser aplicado às políticas de avaliação de risco.

141. Para que todo o potencial dos algoritmos seja explorado, respeitando os princípios de proteção de dados, o princípio da precaução deve ser aplicado por meio de políticas preventivas, que devem ser postas em prática para combater os riscos potenciais associados ao uso dos dados processados por esses algoritmos e o impacto do seu uso nos indivíduos e na sociedade em geral.

142. O princípio da legalidade no tratamento de dados pessoais e a obrigação de prevenir ou minimizar o impacto do processamento de dados sobre os direitos e as liberdades fundamentais dos titulares dos dados devem induzir uma avaliação prévia dos riscos. Isto deverá permitir que sejam implementadas medidas adequadas, especialmente durante a fase de projeto, para mitigar os riscos identificados.

143. Como os dados pessoais devem ser processados para fins específicos e legítimos, não podem ser utilizados de forma incompatível com os efeitos e não devem ser processados de forma que a pessoa em causa possa considerar inesperado, inadequado ou questionável (princípio da lealdade). A questão da reutilização de dados pessoais, tornando-os amplamente acessíveis, portanto, precisa ser tratada com a máxima cautela.

144. O design dos métodos de processamento de dados usados pelos algoritmos deve minimizar a presença de dados redundantes ou marginais e evitar potencial viés oculto e qualquer risco de discriminação ou impacto negativo sobre os direitos e liberdades fundamentais dos titulares dos dados.

145. Quando é utilizada inteligência artificial, os direitos dos titulares de dados são de especial importância e o controle que cada um de nós deve ter sobre nossas informações pessoais implica que os seguintes direitos podem ser exercidos: o direito de não estar sujeito a uma decisão automatizada que afeta significativamente a pessoa em causa sem que as suas opiniões sejam tidas em conta o direito de obter informações sobre o raciocínio subjacente ao processamento de dados realizado pelos algoritmos, o direito de se opor a tal processamento e o direito a um recurso legal.

9. O potencial e as limitações das ferramentas de justiça preditiva

O termo “justiça preditiva” deve ser descartado porque é ambíguo e enganoso. Estas ferramentas baseiam-se em métodos de análise da jurisprudência, utilizando métodos que não reproduzem de forma alguma o raciocínio jurídico, mas podem tentar descrevê-lo. Vieses analíticos, se não puderem ser totalmente eliminados, devem ser identificados. O processo de design e o uso da ferramenta devem se encaixar em uma estrutura ética perfeitamente clara.

146. Na seção 3, já destacamos a ambiguidade e a falácia do conceito de justiça preditiva e como ela opera uma lenta mudança na mente coletiva, levando-nos a acreditar que máquinas, desprovidas de qualquer emoção, um dia serão mais capazes de tornar o ato de julgar mais confiável. Essas promessas precisam ser examinadas de maneira objetiva e científica, com base em fundamentos sólidos de pesquisa fundamental, para detectar possíveis limitações. Neste contexto, deve notar-se que os riscos de interpretações distorcidas do sentido das decisões judiciais são extremamente altos quando baseados apenas na modelagem estatística. Esta observação é agravada pela falta de compreensão precisa das ligações entre os dados e a presença óbvia de falsas correlações que não podem ser discernidas em grandes massas de dados.

147. Além disso, a neutralidade dos algoritmos é um mito, na medida em que seus criadores transferem, conscientemente ou não, seus próprios sistemas de valores para eles. O filósofo Eric Sadin observou que, por trás de sua eficiente e impessoal fachada, os sistemas algorítmicos refletem imperceptivelmente as intenções de seus criadores ou patrocinadores, induzindo um poder funcional e assimétrico às vidas de outras pessoas. Da mesma forma, o pesquisador Aurélien Grosdidier considera que um algoritmo, por si só, não é capaz de nada mais do que nos permitir – na melhor das hipóteses – compreender parte das intenções do designer e ampliar o questionamento para toda cadeia de processamento de informações (intenção do designer, realização do código de computador, execução do código de computador e contexto de execução e, em seguida, manutenção). Esta observação também é compartilhada pelo criminologista Aleš Završnik, que aponta como as etapas de construção e interpretação de algoritmos são realizadas pelo homem, para o homem, e não podem escapar de erros, preconceitos, valores, interesses humanos e uma representação humana do mundo, seja qual for a maneira como eles são projetados.

148. Apesar dessas limitações significativas, deveríamos ignorar a contribuição de uma tecnologia com poder inigualável? Os matemáticos C. S. Calude e G. Longo enfatizam em seu estudo sobre big data que o escopo negativo ou limitante de seus resultados, como acontece com frequência, não destrói a ciência dos dados, mas abre o caminho para uma reflexão maior, incluindo o desafio de um novo método científico mais amplo, capaz de incorporar novos instrumentos algorítmicos e ferramentas clássicas, acompanhando o processamento com uma avaliação rigorosa das evidências. Como destacado na introdução, o uso de IA provavelmente oferecerá apoio extremamente significativo para profissionais, incluindo juízes e advogados, mas também para o público em geral, especialmente se um dia eles possibilitarem a construção de ferramentas inigualáveis de pesquisa e de análise documental em questões legislativas, regulatórias, jurisprudenciais e doutrinárias e criarem vínculos dinâmicos entre todas essas fontes. Mas esse tipo de aplicação vai além do escopo deste artigo, uma vez que não é uma questão de “prever” o resultado de um litígio, mas analisar a jurisprudência num determinado tempo e espaço.

149. Conforme discutido na seção 6, sujeita à representatividade de amostras processadas, a IA ajudou a criar escalas muito mais precisas da média em vários campos (pensão alimentícia, benefícios compensatórios, compensação por lesões corporais, indenizações por demissão, etc.). Essas escalas, baseadas mais em um consenso do que numa análise média do que já existe, já fornecem apoio significativo para a tomada de decisões e orientações, sem poder substituir a lei em si. Como mencionado anteriormente, o risco é que, na ausência de uma representação estatística da realidade ou de ser capaz de prever qualquer coisa, os resultados do software de justiça preditiva sejam definidos como padrões sem validação pelo sistema jurídico e em contradição com o mesmo.

150. Por fim, consideremos a ideia de poder dar um passo atrás em sistemas preditivos. Em vez de bloquear usuários em uma probabilidade (ou conjunto de probabilidades), a ideia seria permitir-lhes navegar através das correlações que levaram o sistema a propor a sua avaliação e a poder selecionar outros conceitos ou grupos mais relevantes de palavras ou excluir falsas correlações. Para usar o exemplo da UCL, isso consistiria em propor uma representação gráfica dos diferentes termos retidos pelo sistema (com suas respectivas ponderações) para descobrir se houve uma violação (ou não violação) e para autorizar outros caminhos, propondo a seleção de outros termos ou grupos lexicais.

151. Por mais ousada e sedutora que seja essa proposta, ela pressupõe que os próprios profissionais (juízes, advogados, universidades) participem coletivamente para testar sua viabilidade e não permitam que operadores privados sozinhos, salvo alguns cientistas não controlados, projetem software e modos de raciocínios ou cálculos bloqueados.

152. As promessas ambiciosas (e não cumpridas) de algumas empresas jurídicas de tecnologia não devem esconder o imenso potencial das tecnologias e a necessidade de aplicações adaptadas e construídas diretamente com pesquisas científicas e acadêmicas, bem como com todos os profissionais do direito, como magistrados, juristas, advogados, notários, oficiais de justiça e especialistas de campo. Certo número de medidas parece ser capaz de tirar o máximo proveito dessas novas ferramentas através de aplicações adaptadas e concebidas em associação direta com pesquisadores e todos os profissionais da área jurídica, incluindo juízes, promotores, escriturários, advogados, notários, oficiais de justiça e especialistas na área.

153. Neste contexto dinâmico, parece essencial, em primeiro lugar, não tomar decisões antecipadamente e discutir os riscos e as aplicações práticas destes instrumentos para os sistemas judiciais, bem como testá-los. Um sistema judicial de acordo com o seu tempo seria aquele capaz de estabelecer, administrar e garantir uma ciberética genuína para setores público e privado, insistindo na total transparência e lealdade dos algoritmos, que podem, um dia, contribuir para a tomada da decisão judicial.

10. A necessidade de um debate público aprofundado sobre estas ferramentas antes da implementação de políticas públicas para o seu desenvolvimento. A necessidade urgente da ciberética para o desenvolvimento de algoritmos de inteligência artificial, respeitando os direitos fundamentais

Os desafios de integrar essas ferramentas num processo de tomada de decisão judicial justifica a simplificação dos conceitos para o grupo alvo envolvido. Uma estrutura ética deve ser estabelecida para promover o rápido desenvolvimento de uma forma de IA que integre, em seus processos de desenvolvimento, mecanismos que previnam possíveis vieses e discriminação.

10.1. A importância de debater, testar e continuamente revisar a aplicação dessas ferramentas antes da implementação de políticas públicas

154. As questões relacionadas com a implementação de ferramentas de justiça preditiva são tão numerosas e multifacetadas que exigem uma abordagem equilibrada por parte dos gestores públicos.

155. Em primeiro lugar, é essencial realizar um debate público em torno dessas questões, reunindo os desenvolvedores das ferramentas e os profissionais da justiça. Os Conselhos judiciais, associações profissionais de juízes e ordens de advogados podem, sem dúvida, contribuir para isso e ajudar a identificar oportunidades e aspectos mais controversos. Além disso, as escolas de magistrados e advogados podem desempenhar um papel fundamental na conscientização dos profissionais de justiça em torno dessas questões, para que possam entender melhor e contribuir de forma prática para os desenvolvimentos em curso.

156. É também essencial realizar investigação sobre as aplicações propostas e testá-las, tanto para compreender o seu potencial e suas fraquezas, mas também desenvolvê-las ainda mais e adaptá-las às nossas necessidades. O direito de examinar os componentes e características dos instrumentos propostos pelo setor privado (ou aqueles desenvolvidos por organizações independentes e institutos públicos, uma solução que deve ser encorajada) parece igualmente importante para que o serviço de justiça possa efetivamente cumprir sua missão. Uma avaliação rigorosa dos resultados dos testes deve

ser realizada antes de uma implantação e integração mais ampla nas políticas públicas. Também parece fortemente aconselhável avaliar regularmente o impacto dessas ferramentas no trabalho dos profissionais da justiça.

10.2. O estabelecimento de um quadro ético

157. Em primeiro lugar, a simples adoção de um quadro legislativo ou regulamentar da IA não é suficiente num contexto digital, que é inerentemente transnacional em seu escopo. Por outro lado, a atenção escrupulosa à natureza e qualidade dos dados abertos e sua qualidade, é possível que minimize os riscos de referências cruzadas inadequadas e reforce a relevância dos resultados do processamento automatizado. No que diz respeito aos nomes dos profissionais, uma simples precaução seria proibir a sua disseminação pública em bases de dados estruturadas e aproximadas no que diz respeito aos riscos de utilização indevida. Não seria uma questão de limitar o acesso à informação já processada (por exemplo, a composição de um painel de juízes), mas de filtrar os dados brutos disponibilizados gratuitamente. Em resumo, distinguir o acesso à informação, do acesso à base de dados, que podem ser manipulados à vontade.

158. Os pesquisadores Buttarelli e Marr enfatizaram como os dados grandes (big data) precisam de um grande controle e proteção. Outros pesquisadores (Pasquale e Morozov) salientaram a necessidade de estabelecer procedimentos transparentes para a implantação de big data e, de um modo mais geral, da IA no domínio judicial, porque as soluções propostas nunca podem representar a vida na sua complexidade.

159. O desenvolvimento de regras de ciberética para orientar a atividade das partes interessadas no setor e promover os princípios de transparência, lealdade e objetividade da ferramenta, são essenciais. O acompanhamento regular por peritos independentes deve assegurar que os condutores de inteligências artificiais utilizadas para auxiliar os juízes em suas decisões não são tendenciosos. Não é inadequado antecipar a implementação, discreta ou não, de sistemas de referência pagos (com base no modelo de publicidade do mecanismo de pesquisa do Google). Alguns operadores atribuam menos peso às decisões que lhes são desfavoráveis. Estas regras terão um papel fundamental no aumento da confiança dos cidadãos nos seus sistemas judiciários.

160. A este respeito, a qualidade dos melhores sistemas poderia ser reconhecida pela atribuição de um rótulo ou certificação. Em particular, eles devem garantir total transparência e perfeita lealdade na forma como a informação é processada, tanto para profissionais como para cidadãos, para evitar a repetição de erros tal como o algoritmo COMPAS, acima mencionado. Profissionais de justiça devem estar estreitamente envolvidos para poder avaliar adequadamente os riscos e o impacto estas aplicações nos sistemas judiciais.

161. Atualmente, todos os especialistas envolvidos no desenvolvimento da IA, incluindo pesquisadores, engenheiros e desenvolvedores de computadores, têm responsabilidades sem precedentes. Seu trabalho pode ser acompanhado por um fortalecimento ainda maior das “humanidades”. O exemplo de algumas escolas inovadoras de desenvolvedores de computadores mostra que por trás da vontade de “hackear o sistema” está se escondendo, segundo alguns observadores, na realidade um pragmatismo desprovido de qualquer contextualização da responsabilidade que agora paira sobre os técnicos com poderes quase demiurgo. O juramento de Hipócrates certamente tem seus limites na medicina, mas ritualiza a tomada de responsabilidades e oferece um quadro ético.

162. Finalmente, a ciberética deve ser acompanhada de treinamento amplo das partes interessadas, desde projetistas de algoritmos até dos usuários. Novas humanidades, transdisciplinar, devem ser compartilhadas entre todos para que a IA se torne o vetor de desenvolvimentos positivos para a humanidade e fortaleça as garantias da democracia e do Estado de direito.

ANEXO II – QUAIS OS USOS DA IA NOS SISTEMAS JUDICIAIS EUROPEUS?

Este anexo da Carta analisa diferentes utilizações da IA nos sistemas europeus e encoraja, em diferentes graus, a sua aplicação à luz dos princípios e valores estabelecidos na Carta Ética.

O uso do aprendizado de máquina para constituir mecanismos de busca para o aprimoramento da jurisprudência é uma oportunidade para ser assumida por todos os profissionais da área jurídica. Aplicações adicionais (elaboração de escalas, apoio a medidas alternativas de resolução de litígios, etc.) devem ser consideradas, porém com precauções (qualidade da fonte de dados em particular e não tratamento em massa de todo o litígio em causa). Outras aplicações (“justiça preditiva”) devem ser atribuídas ao campo de pesquisa e desenvolvimento (em consulta com profissionais da área jurídica, a fim de garantir que eles estão totalmente de acordo com as necessidades reais) antes de considerar o uso em escala na esfera pública.

Em matéria penal, essa é uma questão muito delicada, mas não deve ser ignorada. Tendo em conta as muitas questões existentes quanto à sua compatibilidade com um número de direitos fundamentais, o uso de potenciais algoritmos de cálculo dos riscos de reincidência de um indivíduo levados à justiça devem ser considerados com reservas extremas. Por outro lado, o processamento de dados quantitativos globais para a prevenção da delinquência é um caminho a ser explorado com essas novas técnicas, levando em consideração vieses conhecidos (efeitos performativos, qualidade dos dados, etc). Da mesma forma, o uso de algoritmos para formar uma melhor ligação entre o tipo de serviço comunitário disponível e a personalidade de um indivíduo pode ser um fator de eficiência para esse tipo de medida.

Utiliza-se para encorajar

- Aprimoramento da jurisprudência: as técnicas de aprendizado de máquina foram cada vez mais utilizadas no campo do processamento de linguagem natural nos últimos anos (o que inclui esforços iniciais para entender a linguagem natural) e constituem um trunfo considerável para a construção de soluções complementares de pesquisa de palavras-chave atuais ou pesquisa de “texto completo”. Essas ferramentas podem ligar várias fontes (por exemplo, constituições e convenções, leis, jurisprudência e doutrina).

Técnicas de visualização de dados podem ilustrar o resultado da pesquisa.

- Acesso ao direito: sem substituir a intervenção humana, os chatbots poderiam ser configurados para facilitar o acesso às várias fontes de informação existentes usando linguagem natural. Modelos de documentos (aplicações judiciais, contratos de locação, etc.) também podem ser gerados on-line.
- Criação de novas ferramentas estratégicas: o uso da ciência de dados e inteligência artificial sobre dados de atividades judiciais podem ajudar a melhorar a eficiência da justiça, tornando possível, por exemplo, levar a cabo avaliações quantitativas e qualitativas e fazer projeções (por ex. recursos humanos e orçamentais futuros). Indicadores Chave de Performance poderiam ser elaborados nesta base. Recomenda-se que os profissionais da área jurídica, especialmente os juízes, sejam envolvidos na implementação dessas ferramentas.

Possíveis usos, exigindo precauções metodológicas

- Ajuda na elaboração de escalas em certas disputas civis: a análise da massa total de decisões judiciais estatisticamente não faz muito sentido se não identifica todos os fatores causais (explícitos e implícitos nas decisões). saber que a média de indenização compensatória numa determinada área geográfica é maior do que em outro é explicável não por causa do comportamento dos juízes, mas dada a característica do território. A Aprendizagem de Máquina pode, portanto, ser útil para identificar decisões (ver aprimoramento da jurisprudência acima), mas o processamento automatizado de dados por si só não pode produzir informações significativas. A construção de uma amostra relevante de decisões a serem processadas (por exemplo, técnica de pesquisa) é um pré-requisito.
- Apoio à medidas alternativas de resolução de litígios em matéria civil: em alguns países europeus, as ferramentas de "justiça preditiva" são usadas por companhias de seguros para avaliar as chances de sucesso de uma disputa e orientar o litigante para outro método de resolução de litígios, quando se percebe que há pouca chance de sucesso. Além disso, alguns sistemas no exterior oferecem valores de compensação sem transparência real das regras de cálculo. Contudo, a avaliação destes sistemas não pode

ser considerada como imparcial e confiável (veja a seção técnicas de aprendizado de máquina). Decisões são tomadas sobre um cidadão usando bases truncadas. Em outros casos, um litigante pode ser aconselhado, por meio de um agente virtual (chatbot), para optar por uma disputa alternativa após uma análise preliminar dos critérios informados pelo próprio litigante, quando ele for a um tribunal ou pesquisar na web uma informação legal. Este agente relacional pode, se necessário, recomendar ao litigante para procurar o conselho de um advogado ou um serviço de mediação. Em todos estes casos, a presença de um terceiro treinado (mediador baseando-se não apenas em técnicas, mas talvez em escalas calculadas, ou um advogado) parece ser a solução mais apropriada nesta fase.

- Resolução de litígios on-line: quando se desloca para uma plataforma de resolução de litígios on-line, o litigante deve ser informado de maneira clara e compreensível se sua disputa é tratada de maneira totalmente automatizada ou com o envolvimento de mediador ou árbitro. Além disso, as informações fornecidas aos litigantes devem ser honestas e deve evitar dar-lhes a impressão de que um tribunal está envolvido (neste contexto, o termo “tribunal on-line” é frequentemente usado para este tipo de plataforma, enquanto que tecnicamente a sua finalidade é fornecer serviços alternativos de resolução de litígios). Estes são dois fatores essenciais para permitir que os litigantes façam uma escolha informada e decida opor-se e recorrer a um verdadeiro tribunal na acepção do artigo 6º da CEDH. Além disso, tendo em conta os requisitos dos artigos 6º e 13º da CEDH, deve-se sempre considerar o controle sobre o processo de resolução de disputas on-line, e seu resultado, pelos tribunais do Estado, especialmente quando o litigante consentiu na resolução de disputas on-line de forma totalmente automatizada.
- O uso de algoritmos na investigação criminal para identificar onde as infrações penais estão sendo cometidas: este tipo de aplicação poderia dizer respeito não só à polícia, mas também ao Ministério Público no âmbito da prevenção de delinquência. Sistemas têm sido usados nos Estados Unidos para orientar as patrulhas policiais em tempo real para possíveis locais onde as ofensas estão sendo cometidas. No entanto, este tipo de abordagem quantitativa pode gerar um forte “efeito performativo” (em um determinado local, há uma chance maior de descobrir uma ofensa e isso

então reforça o sistema). Análise criminal através de abordagens acopladas entre sistemas de informação geográfica (GIS) e massas de dados dos procedimentos poderia ser melhor compartilhada com o Ministério Público e certamente se beneficiar de uma contribuição significativa de aprendizado de máquina. Unidades de combate à lavagem de dinheiro já usam sistemas “preditivos” para identificar fluxos financeiros suspeitos, mas no caso de informações quantitativas (financeiras), as máquinas são mais capazes de produzir resultados. Os pesquisadores também devem ter melhor acesso a esses dados para produzir estudos relevantes para os julgadores.

Usos a serem considerados após estudos científicos adicionais

- Perfil do juiz: quantificar a atividade de um juiz revelará menos sobre quaisquer possíveis vieses do que sobre quaisquer fatores externos que influenciam decisões. O juiz em si não é a razão pela qual um juízo em atividade numa área empobrecida não produz os mesmos resultados que em outro território, qualquer que seja sua personalidade. Quando a decisão é proferida de forma colegial e sem a possibilidade de um juiz expressar uma opinião divergente, é inútil fazer um perfil de cada um dos juizes da câmara. Por outro lado, oferecer aos juizes mais avaliação quantitativa e qualitativa detalhada de suas atividades, graças a novas ferramentas, mas com um objetivo meramente informativo de ajudar na tomada de decisão e para seu uso exclusivo, poderia ser incentivado.
- Antecipar decisões judiciais: processamento estatístico de grupos lexicais só revela a frequência do uso de certos grupos de palavras, mas não identifica as razões reais de uma decisão, porque não é o resultado de um raciocínio jurídico (ver o estudo realizado sobre as decisões da CEDH pela University College of London, que produziu melhores resultados a análise dos fatos do que sobre a análise da lei). Sistemas híbridos, fundados sobre a construção de modelos matemáticos que deveriam representar a diversidade de raciocínio dos juizes, não são mais eficazes porque se limitam a amostra de dados que eles processaram e precisam ser completamente reescritos em caso de alteração da lei ou reversão da jurisprudência.

Usos a considerar com reservas mais extremas

- Uso de algoritmos em questões criminais para o perfil de indivíduos: experimentos em outros países (COMPAS nos Estados Unidos e HART no Reino Unido) foram criticados por ONGs (veja trabalho da ProPublica nos Estados Unidos e do Big Brother Watch no REINO UNIDO). Devido às limitações da metodologia utilizada, esta pura abordagem estatística levou a um resultado errado: a descoberta de que Indivíduos afro-americanos são mais frequentemente envolvidos em atos criminosos levou a um fator de risco mais alto para toda a população afro-americana. Então, mesmo para fatos menores, esses sistemas pesaram negativamente os afro-americanos acusados e levaram a um aumento injusto no quantum de suas sentenças. Essa abordagem, que tem efeitos discriminatórios e determinísticos, deve ser substituída por uma que respeite mais as normas europeias em matéria de sanções penais e que deve oferecer ao indivíduo a possibilidade de reabilitação e reintegração. Se os sistemas algorítmicos conseguirem ajudar a melhorar o agrupamento de informações para serviços de estágio, por exemplo, é possível que as informações relevantes sejam coletadas mais rapidamente para posterior processamento humano, tais iniciativas podem ser úteis (especialmente para procedimentos rápidos de julgamento). Qualquer outro uso é propenso a preconceitos que entrarão em conflito com determinados princípios fundamentais supranacionais.
- Norma baseada na quantidade: não é apenas uma questão de produzir escalas, que poderia ser legítimo, mas de fornecer a cada juiz o conteúdo das decisões produzidas por todos os outros juízes e bloquear sua escolha futura na massa desses “precedentes”. Esta abordagem deve ser rejeitada porque este grande número não pode substituir a lei, pelas razões fornecidas acima uma abordagem baseada na quantidade não é o caminho. O estudo do CEPEJ também destacou os perigos da cristalização da jurisprudência e os efeitos potencialmente negativos sobre a imparcialidade e independência dos juízes.

ANEXO III: GLOSSÁRIO

Este glossário fornece uma definição dos termos utilizados pela Carta Ética, objeto do estudo. Foi utilizada uma definição restrita. O documento deve ser lido e compreendido à luz dessas definições.

A

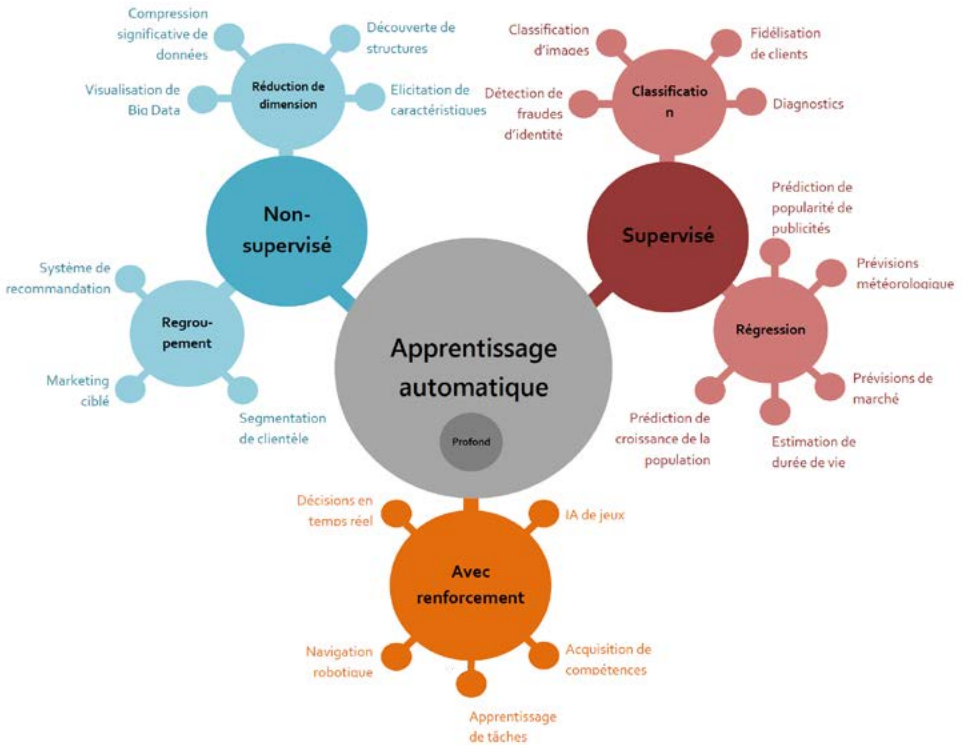
ALGORITMO Sequência finita de regras formais (operações e instruções lógicas) que permitem obter um resultado a partir da entrada inicial de informação. Esta sequência pode ser parte de um processo de execução automatizado e desenhar modelos projetados através do aprendizado de máquina.

ANONIMIZAÇÃO Método para processar dados pessoais a fim de impedir irreversivelmente a identificação de uma pessoa singular ou coletiva. A anonimização implica, por conseguinte, que já não existe qualquer ligação possível entre a informação em causa e a pessoa a quem se refere. A identificação torna-se completamente impossível.⁸⁴

APRENDIZAGEM DE MÁQUINAS (MACHINE LEARNING) A aprendizagem de máquina torna possível construir um modelo matemático a partir dos dados, incorporando um grande número de variáveis que não são conhecidas de antemão. Os parâmetros são configurados gradualmente durante a fase de aprendizagem, que usa conjuntos de dados de treinamento para encontrar e classificar links. Os diferentes métodos de aprendizado de máquina são escolhidos pelos projetistas dependendo da natureza das tarefas a serem concluídas (agrupamento). Os métodos são geralmente classificados em três categorias: aprendizagem supervisionada (humana), aprendizagem não supervisionada e aprendizagem de reforço. Essas três categorias agrupam diferentes métodos, incluindo redes neurais, aprendizagem, etc.

⁸⁴ Grupo de trabalho sobre o artigo 29 do parecer 05/2014 sobre as técnicas de anonimização. Veja também o N° 26 do Regulamento (UE) 2016/679 do Parlamento e Conselho Europeu, de 27 de abril de 2016.

O gráfico abaixo ilustra as diferentes categorias de aprendizado de máquina:



APRENDIZAGEM PROFUNDA (DEEP LEARNING) Veja Machine Learning e Neurons

B

BANCO DE DADOS (DATABASE) Um banco de dados é um “contêiner” que armazena dados como números, datas ou palavras, que podem ser reprocessadas usando um computador para produzir informações, por exemplo, coleta e classificação de números e nomes para formar um diretório.

BIG DATA (metadados, grandes conjuntos de dados) O termo big data se refere a grandes conjuntos de dados de fontes mistas (por exemplo, dados abertos, dados proprietários e dados comprados comercialmente). Para dados derivados de atividade judicial, big data poderia ser a combinação de dados estatísticos, registros de conexões de software de negócios (logs de aplicativos), bancos de dados de decisões judiciais etc.

C

CIÊNCIA DE DADOS (DATA SCIENCE) Um grande campo que agrupa matemática, estatística, probabilidades, processamento de dados e visualização de dados, a fim de obter entendimento a partir de um conjunto misto de dados (imagens, som, texto, dados do genoma, links entre redes sociais, medições físicas, etc.).

Métodos e ferramentas derivadas da inteligência artificial se enquadram nessa categoria.

CHATBOT (agente de conversação) Agente que conversa com seu usuário (por exemplo, robôs de empatia usados para ajudar aqueles que estão doentes ou serviços automatizados de conversação nas relações com os clientes).⁸⁵

D

DADOS (DATA) Representação de informações para processamento. Por fim, às vezes há confusão entre os dados (estritamente falando dados abertos) e seus métodos de processamento (aprendizado de máquina, ciência de dados) para diferentes propósitos (mecanismos de busca, assistência na elaboração de documentos, análise de tendências de decisões, prevendo decisões judiciais, etc.).

DADOS PESSOAIS (PERSONAL DATA) Qualquer informação relativa a identificação de uma pessoa natural (a “pessoa interessada”), direta ou indiretamente.

I

INTELIGÊNCIA ARTIFICIAL (IA) Um conjunto de métodos científicos, teorias e técnicas cujo objetivo é reproduzir, por uma máquina, as habilidades cognitivas dos seres humanos. Desenvolvimentos atuais objetivam que as máquinas executem tarefas complexas anteriormente realizadas por seres humanos.

No entanto, o termo inteligência artificial é criticado por especialistas que distinguem entre IAs “fortes” e “fracas” ou “moderadas”. Alguns especialistas argumentam que IAs “fortes” exigiriam avanços significativos na pesquisa básica,

⁸⁵ Relatório da CNIL de dezembro de 2017: Como os humanos podem manter a vantagem? As questões éticas criadas por algoritmos e inteligência artificial

e não apenas melhorias simples no desempenho dos sistemas existentes, para poder modelar o mundo como um todo.

As ferramentas identificadas neste documento são desenvolvidas usando métodos de aprendizado de máquina, i. e. IAs “fracos”.

J

JUSTIÇA PREDITIVA A justiça preditiva é a análise de grandes quantidades de decisões judiciais por meio de tecnologias de inteligência artificial para fazer previsões do resultado de certos tipos de disputas especializadas.

L

LEGAL TECH Empresas que utilizam tecnologia da informação no campo do direito, a fim de oferecer serviços jurídicos inovadores. Essas empresas são start-ups especializadas em direito.

M

METADATA Dados que permitem definir, contextualizar ou descrever outros dados. Na maioria dos usos do computador, o prefixo meta significa “referência definição ou descrição”. Metadados sintetizam informações básicas sobre dados, facilitam a busca e manipulação de instâncias de dados particulares.

MINERAÇÃO DE DADOS (DATA MINING) A mineração de dados permite analisar um grande volume de dados e destacar modelos, correlações e tendências.

P

PROCESSAMENTO DE DADOS PESSOAIS (PROCESSING OF PERSONAL DATA) significa qualquer operação ou conjunto de operações realizadas em dados pessoais, como coleta, armazenamento, preservação, alteração, recuperação, divulgação, disponibilização, apagamento ou destruição.

PROFILING técnica automatizada de processamento de dados que consiste em aplicar um “perfil” a uma pessoa singular, em particular, para tomar

decisões sobre ele ou ela ou para analisar ou prever preferências pessoais, comportamentos e atitudes.

R

REDES NEURAIS ARTIFICIAIS (NEURONS / NEURAL NETWORK) Redes neurais são sistemas de computação vagamente inspirado pelas redes neurais biológicas que constituem cérebros. Tais sistemas “aprendem” a realizar tarefas considerando exemplos, geralmente sem ser programado com regras específicas de tarefas. Por exemplo, no reconhecimento de imagem, eles podem aprender a identificar imagens que contenham gatos analisando exemplos de imagens que foram rotuladas manualmente como “gato” ou “nenhum gato” e usando os resultados para identificar gatos em outras imagens. Eles fazem isso sem qualquer conhecimento prévio sobre gatos, por exemplo, que eles têm pelos, caudas, bigodes e rostos de gato. Em vez disso, eles geram automaticamente características do material de aprendizagem que processam.

Uma RNA é baseada em uma coleção de unidades conectadas ou nós chamados artificiais neurônios que modelam vagamente os neurônios em um cérebro biológico. Cada conexão, como as sinapses de um cérebro biológico, pode transmitir um sinal de um neurônio artificial para outro. Um neurônio artificial que recebe um sinal pode processá-lo e, em seguida, sinalizar neurônios artificiais adicionais ligados a ele.

O objetivo original da abordagem de RNA era resolver problemas da mesma maneira que um cérebro humano faria. No entanto, com o tempo, a atenção passou a executar tarefas específicas, levando a desvios da biologia. A Rede Neural Artificial tem sido usada em uma variedade de tarefas, incluindo visão computacional, reconhecimento de fala, tradução automática, filtragem de rede social, jogando tabuleiro e vídeo games e diagnóstico médico.

S

SISTEMA ESPECIALISTA Esta é uma das maneiras de obter inteligência artificial. O sistema especialista é uma ferramenta capaz de reproduzir os mecanismos cognitivos de um especialista em um campo particular. Mais precisamente, é um software capaz de responder perguntas, por raciocínio baseado em fatos e

regras conhecidas. Consiste em 3 partes:

- uma base de fatos;
- uma base de regras;
- um motor de inferência.

O mecanismo de inferência é capaz de usar fatos e regras para produzir novos fatos, até ele alcançar a resposta para a pergunta do especialista. A maioria dos sistemas especialistas existentes é baseada em mecanismos lógicos formais (lógica aristotélica) e usa o raciocínio dedutivo.

SOFTWARE DE FONTE ABERTA (OPEN SOURCE SOFTWARE) Software para o qual o código-fonte está disponível para todos. O software pode, portanto, ser usado livremente, modificado e redistribuído.

ANEXO IV: CHECKLIST PARA INTEGRAR OS PRINCÍPIOS DA CARTA NOS SEUS TRATAMENTOS

Para avaliar a compatibilidade do seu método de processamento com a Carta, existe uma escala de auto avaliação disponível para cada um dos princípios listados.

Para cada princípio, marque a caixa que corresponde aos seus métodos de processamento.

A caixa mais à esquerda indica integração completa, a caixa mais à direita indica que não há integração.



A coluna com a maior pontuação indica o nível de compatibilidade do método de processamento com o regulamento.

Esta avaliação é, obviamente, puramente informativa e de forma alguma equivale a qualquer certificação.

CHECHLIST PARA AVALIAR SEUS MÉTODOS DE PROCESSAMENTO

Princípio do respeito pelos direitos fundamentais:

Assegurar que a elaboração e a implementação de ferramentas e serviços de Inteligência Artificial sejam compatíveis com os direitos fundamentais

Lista de verificação para avaliar seus métodos de processamento

Princípio da não discriminação:

Prevenir o desenvolvimento ou intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos

Lista de verificação para avaliar seus métodos de processamento

Princípio da qualidade e segurança:

No processamento de decisões e dados judiciais, utilizar fontes certificadas e dados intangíveis, com modelos elaborados de forma multidisciplinar, num ambiente tecnológico seguro

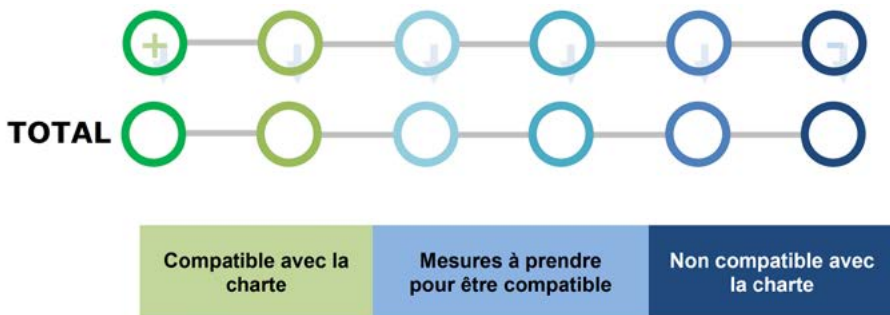
Lista de verificação para avaliar seus métodos de processamento

Princípio da transparência, imparcialidade e justiça:

Tornar os métodos de processamento de dados acessíveis e compreensíveis, autorizar auditorias externas

Lista de verificação para avaliar seus métodos de processamento

Princípio “sob controle do usuário”: impedir uma abordagem prescritiva e garantir que os usuários sejam informados e controlem suas escolhas



**COMISSÃO EUROPEIA PARA A
EFICIÊNCIA DA JUSTIÇA (CEPEJ)**

Carta Ética Europeia sobre o uso da Inteligência Artificial em sistemas judiciais e seu ambiente

A stylized map of Europe is centered in the lower half of the page. The map is composed of various colored regions in shades of blue, green, and brown. Overlaid on the map is a graphic of a document with a white cover and a light blue page, appearing to be turned or held open. The background of the entire page features abstract, flowing shapes in shades of blue and white.

**Charte éthique européenne d'utilisation de
intelligence artificielle dans les systèmes
judiciaires et leur environnement**



Strasbourg, 3 décembre 2018

CEPEJ(2018)14

COMMISSION EUROPEENNE POUR L'EFFICACITE DE LA JUSTICE
(CEPEJ)

**Charte éthique européenne d'utilisation de l'intelligence artificielle
dans les systèmes judiciaires et leur environnement**

adoptée par la CEPEJ lors de sa 31^{ème} réunion plénière
(Strasbourg, 3-4 décembre 2018)



Table des matières

Introduction	4
Les 5 principes de la Charte éthique d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement	5
1. <i>Principe de respect des droits fondamentaux : Assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux</i>	6
2. <i>Principe de non-discrimination : Prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus</i>	7
3. <i>Principe de qualité et sécurité : En ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi disciplinaire, dans un environnement technologique sécurisé</i>	8
4. <i>Principe de transparence, de neutralité et d'intégrité intellectuelle : Rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes</i>	9
5. <i>Principe de maîtrise par l'utilisateur : Bannir une approche prescriptive et permettre à l'usager d'être un acteur éclairé et maître de ses choix</i>	10
Annexe I : Etude approfondie sur l'utilisation de l'IA dans les systèmes judiciaires, notamment les applications d'IA assurant le traitement des décisions et des données judiciaires	11
Introduction	12
1. Etat des lieux de l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires des Etats membres du Conseil de l'Europe	14
2. Etat des lieux des politiques publiques d'open data des décisions judiciaires dans les systèmes judiciaires des Etats membres du Conseil de l'Europe	16
2.1 Définition de l'open data des décisions judiciaires	16
2.2 Etat du développement de l'open data des décisions judiciaires dans les Etats-membres du Conseil de l'Europe et conséquences sur le développement de la jurisprudence	17
2.3 La protection des données à caractère personnel dans les politiques d'ouverture des décisions judiciaires	21
3. Caractéristiques de fonctionnement de l'intelligence artificielle (apprentissage machine) appliquée aux décisions judiciaires	24
3.1 Les fonctionnalités théoriques d'un logiciel de justice dite prédictive	24
3.2 Le fonctionnement concret des intelligences artificielles : des machines statistiques construisant des modèles sur la base du passé	25
4. Les intelligences artificielles peuvent-elles modéliser a priori un raisonnement juridique?	29
5. Les IA peuvent-elles expliquer a posteriori le comportement des juges ?	32
6. Quelles applications pour la justice civile, commerciale et administrative ?	34
6.1 Une nouvelle puissance de calcul pour la constitution de barèmes	35
6.2 Les résolutions de litiges en ligne	35

6.3 Les principales garanties à réaffirmer dans le cadre de la procédure civile, commerciale et administrative	37
7. Les enjeux spécifiques à la justice pénale : prévention des infractions, du risque de récidive et évaluation de la dangerosité	39
7.1 Les outils utilisés par les services d'enquête avant le procès pénal	39
7.2 Les outils durant le procès pénal	40
7.3 Les défis du « prédictif » en matière pénale	42
8. Questions spécifiques relatives à la protection des données personnelles	44
9. Les potentialités et les limites des outils de justice dite prédictive	45
10. La nécessité d'un débat public approfondi autour de ces outils antérieurement à la mise en place des politiques publiques de développement de ceux-ci. L'impérieuse nécessité d'une cyberéthique pour encadrer le développement des algorithmes d'intelligence artificielle dans le respect des droits fondamentaux	47
10.1 L'importance de débattre, tester et continuellement revoir l'application de ces outils antérieurement à la mise en place de politiques publiques	47
10.2 La mise en place d'un cadre éthique	47
Annexe II : Quelles utilisations pour l'IA dans les systèmes judiciaires européens ?	49
Utilisations à encourager	49
Utilisations à envisager avec de fortes précautions méthodologiques	50
Utilisations à envisager au terme de travaux scientifiques complémentaires	51
Utilisations à envisager avec les plus extrêmes réserves	51
Annexe III : Glossaire	52
Annexe IV : Checklist d'intégration des principes de la Charte dans vos traitements	58
Checklist d'évaluation de votre traitement	59

Introduction

Prenant acte de l'importance croissante de l'intelligence artificielle (IA) dans nos sociétés modernes, et des bénéfices escomptables lorsqu'elle sera utilisée pleinement au service de l'efficacité et de la qualité de la justice¹, la CEPEJ adopte solennellement les 5 principes fondamentaux suivants appelés « Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement ».

Cette Charte s'adresse aux acteurs publics et privés en charge de la conception et du déploiement d'outils et de services d'intelligence artificielle s'appuyant notamment sur le traitement des décisions juridictionnelles et des données judiciaires (apprentissage machine ou toutes autres méthodes issues des sciences de données).

Elle concerne également les décideurs publics en charge de l'encadrement législatif ou réglementaire, du développement, de l'audit ou de l'utilisation de tels outils et services.

L'utilisation de tels outils et services dans les systèmes judiciaires a vocation à améliorer l'efficacité et la qualité de la justice et mérite d'être encouragée. Elle doit toutefois se faire de manière responsable, dans le respect des droits fondamentaux des individus énoncés par la Convention européenne des droits de l'homme et par la Convention pour la protection des données à caractère personnel, ainsi que des autres principes fondamentaux énoncés ci-après, qui devraient orienter la définition de politiques publiques de la justice en ce domaine.

Le traitement des décisions juridictionnelles par l'IA est susceptible, selon leurs concepteurs, de contribuer, en matière civile, commerciale et administrative, à améliorer la prévisibilité de l'application de la loi et la cohérence des décisions judiciaires, sous réserve du respect des principes énoncés ci-après. En matière pénale, leur utilisation doit être envisagée avec les plus extrêmes réserves, afin de prévenir des discriminations sur des données sensibles, en conformité avec les garanties du procès équitable.

Qu'ils soient conçus dans le but d'apporter un support à une consultation juridique, une aide à la rédaction ou à la décision ou une orientation des justiciables, il est essentiel que lesdits traitements soient effectués dans des conditions de transparence, de neutralité et de loyauté certifiées par une expertise extérieure à l'opérateur et indépendante.

Application de la Charte

Les principes de la Charte devraient pouvoir faire l'objet d'une application, d'un suivi et d'une évaluation régulière par les acteurs publics et privés, pour améliorer constamment les pratiques.

A cet égard, il est souhaitable qu'un bilan régulier de la mise en œuvre des principes de la Charte soit effectué par ces acteurs, expliquant le cas échéant les raisons de non-réalisation ou de réalisation partielle, accompagné d'un plan d'action pour introduire les mesures nécessaires.

Les autorités indépendantes mentionnées dans la Charte pourraient être chargées d'évaluer périodiquement le niveau d'adhésion aux principes de cette Charte de tous les acteurs et de proposer des améliorations pour l'adapter à l'évolution des technologies et des utilisations de ces technologies.

¹ Pour la définition d'intelligence artificielle, voir le Glossaire en annexe.

Les 5 principes de la Charte éthique d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement

1

PRINCIPE DE RESPECT DES DROITS FONDAMENTAUX : assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux.

2

PRINCIPE DE NON-DISCRIMINATION : prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus.

3

PRINCIPE DE QUALITE et SECURITE : en ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi-disciplinaire, dans un environnement technologique sécurisé.

4

PRINCIPE DE TRANSPARENCE, DE NEUTRALITE ET D'INTEGRITE INTELLECTUELLE : rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes.

5

PRINCIPE DE MAÎTRISE PAR L'UTILISATEUR : bannir une approche prescriptive et permettre à l'utilisateur d'être un acteur éclairé et maître de ses choix.

1

Principe de respect des droits fondamentaux :

assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux

Le traitement de décisions juridictionnelles et de données judiciaires doit répondre à des finalités claires, dans le respect des droits fondamentaux garantis par la Convention européenne des droits de l'homme (CEDH) et de la Convention pour la protection des données à caractère personnel (Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, STE n° 108 telle qu'amendée par le protocole d'amendement STCE n°223).

Lorsque les outils d'intelligence artificielle sont utilisés pour trancher un litige ou en tant qu'outils d'aide à la décision judiciaire ou d'orientation du justiciable, ils ne doivent pas porter atteinte aux garanties du droit d'accès au juge et du procès équitable (égalité des armes et respect du contradictoire).

Ils devraient également être utilisés dans le plein respect des principes de l'Etat de droit et de l'indépendance des juges dans leur processus de décision.

Les approches dites d'« éthique² dès la conception » (*ethical-by-design*) ou de « droits de l'homme dès la conception » (*Human rights by design*) seraient ainsi à privilégier. Elles consistent à intégrer dès la phase de conception et d'apprentissage des règles interdisant de porter atteinte directement ou indirectement aux valeurs fondamentales protégées par les Conventions

² Le choix éthique est réalisé en amont par les concepteurs du programme et n'est donc pas laissé à l'utilisateur.

2

Principe de non-discrimination :

prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus

Conscients de la capacité de ces traitements à révéler des discriminations existantes, par regroupement ou classification de données relatives à des individus ou des groupes d'individus, les acteurs publics et privés doivent s'assurer que ceux-ci ne reproduisent pas ou n'aggravent pas ces discriminations et qu'ils ne conduisent pas à des analyses ou usages déterministes.

Une vigilance particulière doit être accordée tant dans la phase d'élaboration que d'utilisation, d'autant plus lorsque ces traitements se fondent de manière directe ou indirecte sur des données « sensibles ». Peuvent être considérées comme telles l'origine prétendument raciale ou ethnique, les conditions socio-économiques, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale ainsi que les données génétiques, les données biométriques, les données concernant la santé ou concernant la vie sexuelle ou l'orientation sexuelle. Des interventions correctives visant à limiter, et si possible neutraliser ces risques, de même qu'une sensibilisation des acteurs doivent être envisagées lorsque ces discriminations sont constatées.

En revanche, l'utilisation de l'apprentissage automatique et la conduite d'analyses scientifiques pluridisciplinaires visant à lutter contre de telles discriminations devraient être encouragées.

3

Principe de qualité et sécurité :

en ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi-disciplinaire, dans un environnement technologique sécurisé

Les concepteurs de modèles à partir de l'apprentissage automatique doivent pouvoir recueillir de manière large l'expertise des professionnels de la justice concernés (juges, procureurs, avocats, etc.) et des chercheurs / enseignants en droit et en sciences sociales (économistes, sociologues, philosophes par exemple).

La constitution d'équipes de projet mixtes dans des cycles de conception courts pour produire des maquettes fonctionnelles est l'un des modes d'organisation permettant de tirer le meilleur parti de cette pluridisciplinarité.

Les mesures éthiques existantes devraient pouvoir être constamment partagées par ces équipes de projet et enrichies sur la base des retours d'expérience.

Les données dérivant des décisions juridictionnelles et intégrées dans un logiciel qui exécute un algorithme d'apprentissage-machine doivent provenir de sources certifiées et ne doivent pas pouvoir être altérées jusqu'à leur utilisation effective par le mécanisme d'apprentissage. L'ensemble du processus doit ainsi pouvoir être tracé pour s'assurer qu'aucune altération, de nature à modifier le contenu ou le sens de la décision traitée, n'est intervenue.

Les modèles et les algorithmes créés devraient pouvoir être également stockés et exécutés dans des environnements sécurisés, garantissant l'intégrité et l'intangibilité du système.

4

Principe de transparence, de neutralité et d'intégrité intellectuelle :

rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes

Un équilibre doit être trouvé³ entre la propriété intellectuelle de certaines méthodes de traitement et les exigences de transparence (accès à la démarche de conception), de neutralité (absence de biais⁴), de loyauté et d'intégrité intellectuelle (faire primer l'intérêt de la justice), en raison des effets juridiques ou des incidences que ces méthodes peuvent avoir sur les personnes. Il convient de préciser que toute la chaîne de conception et de fonctionnement est concernée par ces mesures puisque la méthodologie de sélection, la qualité et l'organisation des données influent directement sur la phase d'apprentissage.

La totale transparence technique (par exemple *open source* du code et de la documentation) est une première possibilité, limitée parfois par la protection des secrets industriels. Le système pourrait être également explicable dans un langage clair et vulgarisé afin de décrire la manière dont il produit ses résultats, en communiquant par exemple sur la nature des prestations proposées, les outils développés, les performances et les risques d'erreur. Des autorités ou des experts indépendants pourraient être en charge de la certification et de l'audit des traitements ou fournir, en amont, du conseil. Des labels renouvelés régulièrement pourraient être délivrés par les autorités publiques.

³ La proposition faite dans l'étude du MSI-NET du Conseil de l'Europe intitulé « Algorithmes et droits humains », page 46, est intéressante : « La solution consistant à demander la divulgation publique d'algorithmes entiers ou de leur code source est utopique dans ce contexte, les entreprises privées considérant leurs algorithmes comme des logiciels propriétaires stratégiques, et les protégeant donc en conséquence. Il paraît en revanche envisageable d'exiger la publication d'informations partielles mais néanmoins importantes, comme les variables utilisées, les objectifs visés par l'optimisation des algorithmes, les jeux de données d'apprentissage, les valeurs moyennes et les écarts types des résultats obtenus, ou la quantité et le type de données traitées par l'algorithme ». Ou encore les suggestions du rapport intitulé « l'IA pour l'humanité » rédigé par le député Cédric Villani dans le cadre d'une mission confiée par le Premier Ministre de la République française, à page 144 : « les auditeurs pourraient tester l'équité et la loyauté d'un programme (faire ce qui est censé faire), par exemple à travers la soumission de multiples fausses données d'entrée, créer des nombreux profils d'utilisateur suivant des caractéristiques précises »...Mêmes constats dans le rapport de la House of Lords, « AI dans le Royaume Uni : prêts, disposés et capables ? sur l'IA : paras 92, 96-99).

⁴ Les solutions envisagées de manière générale pour garantir la neutralité des algorithmes dans le rapport susmentionné de la House of Lords (§114, 115, 116, 119, 120) sont intéressantes: plus de données diversifiées, plus de diversité et de transdisciplinarité, plus d'audit sur des aspects comme le traitement des données ou la manière dans laquelle la machine est construite.

5

Principe de maîtrise par l'utilisateur :

bannir une approche prescriptive et permettre à l'utilisateur d'être un acteur éclairé et maître de ses choix

L'autonomie de l'utilisateur doit être renforcée et ne pas être restreinte par l'utilisation d'outils et de services d'intelligence artificielle.

Le professionnel de la justice devrait à tout moment pouvoir revenir aux décisions et données judiciaires ayant été utilisées pour produire un résultat et continuer à avoir la possibilité de s'en écarter au vu des spécificités de l'affaire concrète.

Le justiciable devrait être informé dans un langage clair et compréhensible du caractère contraignant ou non des solutions proposées par les outils d'intelligence artificielle, des différentes options possibles ainsi que de son droit à une assistance juridique et à saisir un tribunal. Le justiciable devrait également être informé de manière claire de tout recours à un traitement préalable d'une affaire par de l'intelligence artificielle avant ou pendant un processus judiciaire et être en mesure de s'y opposer pour que son affaire soit entendue directement par un tribunal au sens de l'Article 6 de la CEDH.

D'une manière générale, des actions d'alphabétisation numérique des usagers et des débats impliquant les professionnels de la justice devraient être conduits à l'occasion de la mise en œuvre de tout système d'information s'appuyant sur l'intelligence artificielle.

Annexe I

Etude approfondie sur l'utilisation de l'IA dans les systèmes judiciaires, notamment les applications d'IA assurant le traitement des décisions et des données judiciaires

préparée sur la base des contributions de M. Xavier Ronsin, Premier Président de la Cour d'Appel de Rennes, expert scientifique (France)

et

M. Vasileios Lamos, Chercheur en chef au Service de sciences de l'informatique au University College de Londres, expert scientifique (Royaume-Uni)

avec la contribution de

Mme Agnès Maîtrepiere, magistrat, membre du Comité consultatif de la Convention sur la protection des données à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe (France)

Les experts suivants ont contribué au perfectionnement de l'Etude :

M. Francesco Contini, Chercheur en chef à l'Institut de recherche sur les systèmes judiciaires – Centre National de Recherche (IRSIG-CNR), Bologne (Italie)

M. Francesco De Santis, Professeur de procédures de tutelle des droits de l'homme, Université de Naples (Italie)

M. Jean Lassègue, Philosophe et épistémologue, chargé de recherche au Centre National de Recherche Scientifique (CNRS) et chercheur associé à l'Institut des Hautes Etudes sur la Justice (IHEJ) (France)

Mme Dory Reiling, Juge principal honoraire, experte indépendante en technologies de l'information et en réformes judiciaires (Pays-Bas)

M. Aleš Završnik, Chercheur en chef à l'Institut de criminologie, professeur associé à la Faculté de droit de l'Université de Ljubljana (Slovénie) et chercheur EURIAS 2017-18 au Collegium Helveticum à Zürich (Suisse)

Introduction

1. La vague de transformation numérique de nos sociétés touche de manière encore inégale les systèmes judiciaires des Etats membres du Conseil de l'Europe. De nombreux pays d'Europe semblent déjà avoir développé une réflexion extrêmement avancée avec des applications concrètes - tant sur le plan technologique que sur le plan de l'accompagnement juridique - alors que pour d'autres le sujet apparaît encore comme émergent et se limiterait simplement à pouvoir mettre en œuvre une informatique de gestion efficace.
2. Parmi les technologies à l'œuvre dans cette grande transformation numérique, l'intelligence artificielle (IA) apparaît à la fois comme la plus spectaculaire et la plus redoutable. Aux États-Unis, des « robots avocats » sont déjà à l'œuvre et semblent converser en langage naturel avec les humains. Les *legaltechs*, start-ups spécialisées dans la conception de nouveaux services juridiques, proposent également des nouvelles applications aux professions juridiques, principalement les avocats, services juridiques et assureurs, permettant d'avoir un accès approfondi à l'information juridique et à la jurisprudence. Ces compagnies privées ambitionnent même de prévoir les décisions des juges avec des outils, improprement qualifiés nous le verrons¹, de « justice prédictive ».
3. Un premier examen de ce phénomène conduit toutefois à distinguer ce qui relève des discours commerciaux d'une réalité d'utilisation et de déploiement. L'utilisation concrète et quotidienne de logiciels prédictifs par des juges paraît pour l'instant inexistant dans les Etats membres du Conseil de l'Europe. Des expérimentations locales² ainsi que des travaux universitaires³ visant à explorer les potentialités de ces applications ont pu être menés, mais sans aucun développement généralisé. En effet, l'initiative du développement de ces outils provient très largement du secteur privé, qui a su pour l'instant trouver un début de clientèle auprès des compagnies d'assurance, des avocats et des services juridiques soucieux de réduire l'« aléa judiciaire », c'est-à-dire l'imprévisibilité de la décision judiciaire. Les décideurs publics commencent néanmoins à être de plus en plus sollicités par un secteur privé souhaitant voir ces outils – qui sont parfois des versions « bêta », c'est-à-dire appelés à évoluer ultérieurement - intégrés dans les politiques publiques.
4. En continuité avec la réflexion entamée dans ses *Lignes Directrices sur la conduite du changement sur la cyberjustice*⁴, la CEPEJ se propose d'offrir aux décideurs publics et aux professionnels de la justice des clés pour une meilleure compréhension du phénomène de la « justice prédictive ».
5. Il s'agira premièrement de clarifier les questions relatives à la nature intrinsèque de ces systèmes de traitement de masse de données de jurisprudence, ainsi qu'à leurs limites techniques et théoriques : celles-ci, très peu présentes dans le débat qui s'est installé dans le monde judiciaire, sont pourtant très connues et discutées par les spécialistes de ces technologies (mathématiciens, statisticiens, informaticiens) et méritent d'être relevées.
6. Deuxièmement, le document analysera les avantages et les risques de ces outils. Si leurs promoteurs mettent en relief leurs atouts en matière de transparence, de prévisibilité et d'homogénéisation de la jurisprudence, leurs détracteurs pointent du doigt non seulement les limites mais aussi les biais de raisonnement des logiciels actuellement sur le marché. Les risques inhérents à ces technologies peuvent même aller au-delà de l'acte de juger et toucher des éléments essentiels de fonctionnement de l'État de droit et des systèmes judiciaires, auxquels le Conseil de l'Europe est particulièrement attaché.
7. Il s'agit notamment de principes comme la primauté de la loi : ces outils, dont les effets ne seraient pas seulement incitatifs mais quasiment prescriptifs, dégageraient en effet une nouvelle

¹ Voir en particulier l'encadré du Chapitre 9.

² A titre d'exemple les deux cours d'appel de Douai et de Rennes en France ont expérimenté pendant trois mois en 2017 le bénéfice d'un logiciel présenté comme prédictif auprès d'un panel de juges.

³ Travaux sur un échantillon de 584 décisions de la Cour européenne des droits de l'Homme : Nikolaos Aletras, Dimitrios Tsarapatansis, Daniel Preotjuc-Pietro, Vasileios Lampos, « Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective », publié le 24 octobre 2016, [En ligne], <https://peeri.com/articles/cs-93/>

⁴ Voir notamment le paragraphe 51 du document CEPEJ(2016)13, Lignes directrices sur la conduite du changement vers la Cyberjustice.

forme de normativité, qui tendrait à compléter la loi en encadrant le pouvoir souverain d'appréciation du juge et pourrait aboutir à long terme à une uniformisation des décisions judiciaires sur la base non plus d'un raisonnement individualisé des tribunaux, mais d'un pur calcul statistique par rapport à la moyenne des indemnités allouées antérieurement par d'autres juridictions.

8. Par ailleurs, la compatibilité de ces solutions avec des droits individuels consacrés par la Convention européenne des droits de l'homme (CEDH) est aussi une source de questionnement. Il s'agit notamment du droit à un procès équitable (plus particulièrement du droit au juge naturel établi par la loi, du droit à un tribunal indépendant et impartial, de l'égalité des armes dans les procédures judiciaires) et du droit au respect de la vie privée et familiale, notamment en cas d'insuffisant respect de la protection des données communiquées en *open data*.
9. Tout en étant sensible à ces questions, le document relève l'immense potentiel de l'IA en tant que support au travail des professionnels de la justice. Nul ne doute que certaines applications d'IA en cours de développement ou d'expérimentation, comme par exemple celles visant à améliorer la recherche juridique, peuvent être d'une très grande utilité et rendre le travail judiciaire à la fois plus rapide et plus efficace. Le document met en avant ces exemples positifs et prône de manière générale une utilisation de l'IA qui soit, d'une part, au service des professionnels de la justice et en phase avec leurs besoins et d'autre part, respectueuse des droits individuels garantis par la CEDH et des normes du Conseil de l'Europe, notamment en matière pénale. Loin d'être un simple instrument d'amélioration de l'efficacité des systèmes judiciaires, l'IA devrait renforcer et ne pas diminuer les garanties de l'État de droit tout comme la qualité du service public de la justice.
10. Enfin, le document donne des pistes pour une « gouvernance » de ce phénomène, sous la forme d'une charte éthique qui dégage un certain nombre de principes qui devraient guider le déploiement de l'IA dans les systèmes judiciaires. Il met par ailleurs l'accent sur la nécessité d'une approche prudente s'agissant de l'intégration des outils d'IA dans les politiques publiques. Il est en effet indispensable qu'un débat public implique tous les acteurs – qu'ils soient professionnels de la justice, *legaltechs* ou scientifiques - afin de leur permettre d'une part de partager leur analyse sur la portée et le possible impact de l'introduction des applications d'IA dans les systèmes judiciaires et d'autre part de développer au profit de celles-ci un indispensable cadre éthique. Dans un second temps, ce débat pourrait sortir d'un cadre purement « métier », impliquer les citoyens eux-mêmes et participer ainsi à une certaine forme d'alphabétisation numérique, comme cela a été réalisé par exemple au Canada⁵.

⁵ Déclaration de Montréal, disponible au lien : <https://www.declarationmontreal-iaresponsable.com/demarche>, visité le 16 juillet 2018.

1. Etat des lieux de l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires des Etats membres du Conseil de l'Europe

En 2018, l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires européens demeure pour l'essentiel à l'initiative du secteur privé afin de développer une offre commerciale à destination des assurances, des directions juridiques, des avocats ou des particuliers.

11. L'utilisation de l'IA dans le champ judiciaire paraît assez largement vulgarisée aux Etats-Unis qui ont investi ces outils de manière assez décomplexée, tant en matière civile que pénale⁶.
12. Le recensement des initiatives en matière d'algorithmes d'IA dans les systèmes judiciaires des Etats membres du Conseil de l'Europe est une tâche moins aisée, car la plupart des initiatives proviennent du secteur privé et sont encore peu intégrées dans les politiques publiques.
13. La question de l'utilisation de l'IA dans les systèmes judiciaires a été traitée dans une enquête ligne spécifique, ouverte en avril 2018 de manière large tant aux représentants des Etats membres auprès de la CEPEJ que de la société civile. Le taux de réponse est apparu comme relativement faible et n'a pas permis de dessiner de tendances claires. Certains opérateurs privés n'ont pas semblé sensibles à ce recensement et les membres de la CEPEJ, qui appartiennent pour la plupart à des ministères de la justice ou des conseils supérieurs de justice, ne pas ont en mesure de citer d'autres outils que ceux effectivement utilisés par la sphère publique.
14. En conséquence, l'état des lieux dressé ci-dessous n'est que partiel et ne s'appuie que sur les recherches menées par les experts et le Secrétariat dans la littérature disponible publiquement⁷.
15. Une classification peut être opérée en fonction du service offert. L'intervention de l'IA peut être très variable en fonction des applications. Les grandes familles pourraient être organisées comme suit, à titre illustratif :
 - Moteur de recherche de jurisprudence avancé,
 - Résolution de litiges en ligne,
 - Aide à la rédaction d'actes,
 - Analyse (prédictif, barèmes).
 - Catégorisation des contrats selon différents critères et détection de clauses contractuelles divergentes ou incompatibles,
 - « Chatbots » de renseignement du justiciable ou de support de celui-ci dans sa démarche litigieuse.
16. La Lettonie a déclaré explorer les possibilités de l'apprentissage machine (*machine learning*) pour des besoins d'administration de la justice. Il s'agit essentiellement de traiter les statistiques des juridictions pour dresser de manière prévisionnelle des estimations d'évolution des moyens (humains, financiers) à allouer.
17. D'autres activités menées par des legaltechs n'ont pas été intégrées dans cette classification car ne relevant peu ou pas du tout d'un traitement par de l'intelligence artificielle : certains sites proposent ainsi de l'accès à de l'information juridique, des solutions « cloud », de la signature électronique, ...

⁶ Voir les algorithmes COMPAS ou les outils tels que RAVEL LAW ou l'agent conversationnel ROSS

⁷ Voir la bibliographie sommaire en Annexe IV – des apports substantiels proviennent de Benoît Charpentier (https://media.wix.com/ugd/c21db1_14b04c49ba7f46bf9a5d88581cbda172.pdf) et de Emmanuel Barthe (<http://www.precisement.org/blog/intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html#nb14>), ainsi que de Giuseppe Contissa et Giovanni Sartor (<https://rm.coe.int/giuseppe-contissa-presentation-study-session-artificial-intelligence-2/16808b7586>)

18. Un aperçu non exhaustif de services juridiques utilisant de l'IA de manière opérationnelle apparaît ci-dessous.

Logiciel	Etat	Type
Doctrine.fr	France	Moteur de recherche
Prédicte	France	Analyse (hors pénal)
Case Law Analytics	France	Analyse (hors pénal)
JurisData Analytics (LexisNexis)	France	Moteur de recherche, Analyse (hors pénal)
Luminance	Royaume-Uni	Analyse
Watson/Ross (IBM)	Etats-Unis	Analyse
HART	Royaume-Uni	Analyse (pénal, risque de récidive)
Lex Machina (LexisNexis)	Etats-Unis	Analyse

2. Etat des lieux des politiques publiques d'*open data* des décisions judiciaires dans les systèmes judiciaires des Etats membres du Conseil de l'Europe

La disponibilité de données est une condition essentielle au développement de l'IA, permettant à cette dernière d'effectuer certaines tâches qui étaient autrefois menées par l'homme de manière non automatisée. Plus les données sont disponibles, plus l'IA est à même d'affiner des modèles de prédiction. L'*open data* des décisions judiciaires est donc le préalable à l'activité des legaltechs spécialisés dans les moteurs de recherche ou l'analyse de tendances (justice dite prédictive).

Le traitement de ces données pose un certain nombre d'enjeux, tels que la transformation de la formation de la jurisprudence ou la protection des données personnelles (notamment le nom des professionnels).

19. Les données collectées par l'informatique seraient actuellement le « pétrole » du XXI^{ème} siècle, puisque leur exploitation et leur croisement produiraient une toute nouvelle richesse. Même si cette assertion est contestée par certains acteurs ou auteurs, les succès de l'industrie numérique à l'échelle planétaire durant les dernières décennies confirment l'énorme potentiel de croissance de ce domaine d'activité.
20. La quantification des activités humaines, désormais à une échelle planétaire, ne pouvait laisser échapper les données produites par le secteur public. C'est dans ce cadre que le mouvement d'ouverture des données publiques s'est opéré, en s'appuyant sur des impératifs beaucoup plus anciens, fondateurs de nos Etats de droit.
21. Le développement majeur de ces dernières années provient de l'ouverture de données publiques téléchargeables (*open data*), notamment dans le cadre du « Partenariat pour un gouvernement ouvert » (« PGO » ou *Open Government Partnership* « OGP »). Le PGO est une organisation non gouvernementale réunissant près de 70 Etats membres (dont de nombreux Etats membres du Conseil de l'Europe) avec des représentants de la société civile et des géants du numérique. Cette ouverture vise à améliorer la transparence de l'action publique, à inciter les citoyens à élaborer et évaluer les politiques publiques et à garantir l'intégrité de l'action publique, ainsi que celle de ses agents, par le traitement de masses considérables d'informations structurées en bases de données (*big data*).

2.1 Définition de l'*open data* des décisions judiciaires

22. Redéfinissons tout d'abord la notion d'*open data* avant de traiter de la question de l'impact de l'ouverture des données sur l'activité judiciaire. **Tout d'abord, il existe une fréquente confusion entre accès à l'information et accès aux données (plus précisément, l'accès à l'information en base de données⁸).**
23. En effet, un certain nombre d'informations publiques, nécessitant une large publicité, bénéficient déjà des technologies de l'information pour assurer leur diffusion : ainsi en France, un site gouvernemental *Légifrance.fr* est le premier canal dématérialisé d'informations publiques certifiées, où l'on trouve non seulement des textes législatifs ou réglementaires, mais aussi de la jurisprudence et des nominations à des emplois publics. Cette information centralisée, bien que disponible sur internet, est cependant totalement distincte de l'accès direct à des données organisées et constituées en base, pouvant être téléchargées et traitées par un ordinateur.

⁸ Les données sont des lettres et des chiffres sans signification. L'information est une donnée incluse dans un contexte. C'est le contexte qui donne un sens aux données. Nous pouvons deviner que 2005 est une année, mais sans contexte, nous ne savons pas. Mais dans « en 2005, nous avons réalisé 3 projets » le contexte donne un sens au nombre. Par conséquent, les « données ouvertes » ne sont pas des données au sens de la définition, mais des informations. De même, des grandes données sont aussi de grandes quantités d'informations, et non des données.

24. **L'open data ne concerne donc que la diffusion de données « brutes » en bases informatiques structurées** ; ces données, agglomérées en tout ou partie avec d'autres sources structurées, constituent ce que l'on qualifie des mégadonnées (*big data*). Le Comité Consultatif de la Convention 108 du Conseil de l'Europe définit le *big data* comme « la capacité technologique croissante de collecter, traiter et extraire très rapidement des connaissances nouvelles et prédictives à partir d'un gros volume, d'une grande variété de données et à une vitesse considérable. Sous l'angle de la protection des données, les principaux problèmes ne viennent pas uniquement du volume, de la variété des données traitées et de la vitesse du processus, mais également de l'analyse de ces données au moyen d'un logiciel dans le but d'extraire des connaissances prédictives de nature à orienter un processus décisionnel à l'égard de personnes ou de groupes. Aux fins des présentes lignes directrices, la définition des mégadonnées englobe donc à la fois les données elles-mêmes et le procédé analytique »⁹.
25. **Comme le montre cette définition, on ne doit pas confondre l'ouverture des données avec leurs moyens de traitement.** Nombre de discours sur cette ouverture concernent en réalité les traitements opérés par diverses méthodes avancées que l'on qualifie globalement de *data science*. Justice dite prédictive à l'aide d'intelligences artificielles, moteurs de recherches avancés avec des critères extrêmement précis, robots juristes sont autant d'applications algorithmiques qui sont nourries de données mais qui n'ont rien à voir avec la politique d'ouverture elle-même.
26. Cette politique d'ouverture doit toutefois être interrogée au regard des possibilités qu'elle offre pour des traitements ultérieurs, quelle que soit leur nature. Car si l'on filtre en amont certaines données, par exemple au titre de la confidentialité et du respect de la vie privée, les risques ultérieurs de détournement de finalité semblent pouvoir être réduits.

2.2 Etat du développement de l'open data des décisions judiciaires dans les Etats membres du Conseil de l'Europe et conséquences sur le développement de la jurisprudence

27. Où en sont les Etats membres du Conseil de l'Europe en matière d'*open data* des décisions judiciaires ? Le cycle d'évaluation 2016-2018 de la CEPEJ s'est intéressé pour la première fois à la question de l'*open data* des décisions de justice, source de certains traitements basés sur l'IA. La question de l'anonymisation ou de la pseudonymisation des données, au regard du cadre normatif européen de protection des données personnelles¹⁰, a fait l'objet d'une question spécifique afin d'identifier les mesures particulières mises en œuvre par les Etats membres et les observateurs dans ce domaine particulièrement sensible.

⁹ T-PD(2017)1, Lignes directrices sur la protection des données à caractère personnel à l'ère des mégadonnées.

¹⁰ Règlement général de protection des données (RGPD, Règlement de l'Union européenne 2016/679, et Convention n°108 du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données personnelles.

30. Il existe toutefois une réelle difficulté de mesure de l'impact de l'*open data* sur l'efficacité et la qualité de la justice. Comme indiqué précédemment, l'initiative du réemploi de ces données est en effet essentiellement privée, avec comme cible une clientèle professionnelle (avocats, directions juridiques) et il paraît difficile d'identifier exhaustivement ces réalisations dans le cadre d'une activité seulement intergouvernementale.
31. L'expérience française est représentative des questionnements posés par cette approche et révèle un certain nombre d'enjeux. Il paraît tout d'abord important de souligner que la France a légiféré en 2016 pour imposer un cadre obligatoire de diffusion en *open data* des décisions de justice à ses juridictions.
32. En effet, les articles 20 et 21 de la loi pour une République numérique¹² ont rompu avec la précédente logique de sélection qui disposait¹³ un principe de publicité des « décisions présentant un intérêt particulier rendues par les autres juridictions de l'ordre judiciaire ». Désormais, la loi française pose le principe inverse que tout est publiable, sauf dans des cas précis identifiés par la loi (pour les seules décisions judiciaires) et dans le respect de la vie privée des personnes concernées. Il est toutefois prévu que cette mise à disposition du public des décisions - tant de l'ordre judiciaire qu'administratif - soit précédée d'une analyse du risque de réidentification des personnes.
33. Les gains d'une large diffusion, promus notamment par la Cour de cassation lors de deux colloques en octobre 2016 et en février 2018, sont nombreux : meilleure connaissance de l'activité judiciaire et des tendances jurisprudentielles, gain de qualité pour une justice qui se sait ainsi observée et création d'une toute nouvelle base de référence de faits.
34. Cette volonté théorique de « discipliner » la jurisprudence par le levier numérique soulève toutefois un certain nombre de questions d'ordre général applicables à l'ensemble des pays qui s'engageraient dans la même démarche. Il convient tout d'abord de rappeler certains principes posés par la Cour européenne des droits de l'homme dans des affaires mettant en cause des divergences de jurisprudence internes : la Cour souligne clairement l'exigence de mise en balance entre l'objectif de sécurité juridique (qui implique la prévisibilité de la décision) et l'intérêt de la « vitalité » dans l'interprétation judiciaire¹⁴.
35. Plusieurs éléments doivent être examinés à cet égard, en premier lieu des éléments techniques :
- a) **La collecte de l'intégralité des décisions judiciaires éligibles à une publication n'est pas nécessairement bien organisée entre tous les degrés de juridictions** : certains des « applicatifs métiers » dans les tribunaux européens n'ont pas été conçus pour cela, notamment s'agissant des décisions de première instance, et cette collecte, si elle se veut exhaustive, imposera de créer dans certains pays des nouveaux processus de collation des jugements ;
 - b) **Même si des développements prometteurs sont en cours, il n'existe pas encore de mécanisme automatisé d'anonymisation a posteriori totalement efficace et à même de prévenir tout risque d'identification ou de réidentification** ;
- Et ensuite, également des éléments de fond :
- c) **La transformation de la logique même de production de la jurisprudence est à considérer** ; Quelle est la valeur de la « norme » issue du nombre des décisions rendues sur une matière déterminée ? Cette « norme » ajoute-t-elle à la loi ? Dans ce cas, s'agit-il d'une nouvelle source de droit ?

¹² Cette loi a été adoptée dans le but de mettre le droit français en conformité avec la directive 2013/37/UE du Parlement européen et du Conseil européen du 26 juin 2013, qui modifiait à son tour la directive du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (dite « directive PSI »).

¹³ Article R433-3 du Code de l'organisation judiciaire.

¹⁴ Paroisse gréco-catholique Lupeni et autres c. Roumanie [GC], n. 76943/11, 29/11/2016, § 116.

- d) **Est-il approprié de demander aux magistrats des efforts supplémentaires de motivation pour expliquer les écarts à cette « moyenne »¹⁵ ?** Cette question n'est pas anodine et ne trouve pas de réponse uniforme dans l'ensemble des systèmes judiciaires. Pour s'en tenir au droit civil français¹⁶, « *le juge tranche le litige conformément aux règles de droit qui lui sont applicables* ». Dans ce contexte particulier français, lui demander de justifier sa décision, en intégrant dans la motivation de cette décision, les raisons pour lesquelles il s'écarte de la prétendue tendance majoritaire de la jurisprudence dans la solution du litige (tout en se conformant aux règles de droit qui lui sont applicables), ne reviendrait-il pas à le faire sortir de son office ? Le juge ne trancherait pas le litige conformément aux seules règles de droit, mais aussi conformément à des tendances jurisprudentielles issues des statistiques élaborées par un outil numérique (qui pourrait par ailleurs être biaisé ou élaboré sans contrôle extérieur par un acteur privé, voir infra 6.1 et 9). La question qui se pose ici va bien au-delà d'une spécificité nationale et renvoie d'une manière plus large à la place des outils d'IA dans le processus décisionnel du juge. Elle n'a pas vocation à remettre en cause la place de la jurisprudence dans les sources du droit ou l'autorité du précédent jurisprudentiel, qui sont des principes acquis au niveau européen¹⁷.
- e) **De plus, la " norme " issue du nombre ne tendrait-elle pas à niveler les décisions de justice, sans les ordonner en fonction de la hiérarchie des juridictions dont elles émanent, ce qui méconnaîtrait la portée des décisions des juridictions suprêmes qui dans beaucoup d'Etats européennes sont les garantes de l'interprétation uniforme du droit ?** Quel serait le rapport de " la norme issue du nombre " avec la jurisprudence de celles-ci ? Devenirait-elle, en étant " issue du nombre " un critère déterminant pour ces cours dans la détermination de leur propre jurisprudence, qu'elles devraient à leur tour justifier lorsqu'elles s'écartent de la " moyenne " ?
- f) **Enfin, n'y a-t-il pas un risque de cristallisation de la jurisprudence, selon une logique reproductive ?** Alors que les décisions judiciaires ont vocation à évoluer en fonction de l'évolution du cadre normatif (national, européen ou international), de la jurisprudence de référence (par exemple des juridictions suprêmes, des cours européennes) ou du contexte socio-économique, la « norme » issue du nombre ne deviendrait-elle pas un standard auquel les magistrats seraient incités à se référer sans questionnement, avec un effet induit d'uniformisation excessive des décisions judiciaires ?
36. **Par ailleurs, on peut s'interroger sur les conséquences pour l'usager, qui est censé être le premier bénéficiaire de la transparence de l'action :** est-ce que celui-ci va tirer réellement un avantage du fait de la publication de la totalité des décisions judiciaires sur internet ou ne risque-t-il pas plutôt de se trouver « noyé » par la masse des décisions, sans en tirer forcément des bénéfices sur le plan qualitatif, faute de disposer des connaissances juridiques et de l'appareil critique pour les analyser ?
37. Paradoxalement, on peut se demander si l'*open data* est en mesure de délivrer une information significative puisqu'il permettra certes, par un simple clic sur un lien de télécharger un ensemble considérable de données brutes, mais dont le sens restera totalement abscons pour la majorité des citoyens.
38. Prêter des vertus de transparence et d'information à l'*open data*, c'est donc d'abord espérer que des tiers puissent s'en emparer pour les analyser puis les valoriser. Un modèle économique dans lequel les données publiques de jurisprudence, faisant l'objet d'un travail soutenu par les autorités judiciaires afin de les rendre techniquement « lisibles » par l'IA et conformes aux exigences légales de protection des données personnelles, seraient remises gratuitement au secteur privé et revendues ensuite par celui-ci aux juridictions, aux professionnels et aux citoyens, mérite donc d'être questionné, voire de faire l'objet de critiques.

¹⁵ Eloi Buat-Menard et Paolo Giambiasi, « La mémoire numérique des décisions judiciaires », *Recueil Dalloz*, 2017, p. 1483.

¹⁶ Article 12 du Code de procédure civile français.

¹⁷ A titre d'exemple, la CJUE a affirmé, à propos des conditions d'engagement de la responsabilité d'un Etat membre du fait du contenu d'une décision d'une juridiction nationale suprême, qu'une « violation du droit communautaire est suffisamment caractérisée », « lorsque la décision concernée est intervenue en méconnaissance manifeste de la jurisprudence de la Cour en la matière » (ECJ, affaire C-224/01, Köbler, § 56).

2.3 La protection des données à caractère personnel dans les politiques d'ouverture des décisions judiciaires

2.3.1. Le nom des parties et des témoins

39. Afin de tendre à l'ère numérique vers un juste équilibre entre la nécessaire publicité des décisions de justice et le respect des droits fondamentaux des personnes concernées, qu'elles soient parties ou témoins, leurs noms et leurs adresses ne doivent pas apparaître dans les décisions publiées, notamment au regard du risque de détournement et de réutilisation de ces informations personnelles, ainsi que de la sensibilité particulière des données susceptibles d'être contenues dans certaines décisions¹⁸. L'occultation de ces informations peut se faire de manière systématique au moyen de la mise en œuvre de processus automatisés.
40. D'autres informations « identifiantes » peuvent aussi être occultées (par exemple, numéros de téléphone, adresses électroniques, dates de naissance, prénoms des enfants, prénoms rares, surnoms, lieu-dit). Cette occultation revient, au regard des principes de protection des données personnelles, à une simple pseudonymisation des données, et non à une complète anonymisation. En effet, le volume et la variété d'informations qui sont contenues dans les décisions de justice, alliés à la facilité grandissante de procéder à des croisements avec d'autres bases de données, rendent en pratique impossible de garantir que la personne concernée ne soit pas ré-identifiée. En l'absence d'une telle garantie, ces données ne peuvent être qualifiées d'anonymes et relèvent donc du régime de protection des données personnelles.
41. Le caractère sensible de certaines données personnelles mérite une attention particulière, comme cela est prévu à l'article 6 de la Convention 108. Il en va ainsi des données qui révèlent l'origine ethnique ou raciale, les opinions politiques, l'appartenance syndicale, les convictions religieuses ou autres convictions, ainsi que la santé (physique ou mentale) ou la vie sexuelle, ce qui relève de la sphère la plus intime de la personne concernée.
42. Les décisions de justice peuvent contenir d'autres types de données personnelles, très variées, qui entrent dans cette catégorie des données sensibles. Une autre catégorie de données sensibles intéresse particulièrement l'activité pénale des juridictions. Il s'agit des données concernant les procédures et les condamnations pénales. L'ensemble de ces données sensibles mérite donc une vigilance particulière. Leur diffusion en masse présenterait de sérieux risques de discrimination, de profilage¹⁹ et d'atteinte à la dignité humaine.

2.3.2. Le nom des professionnels, notamment celui des juges

43. De manière évidente, la connaissance de la formation de jugement a toujours été un élément d'anticipation essentiel pour les avocats, pour qui connaître son juge est parfois presque aussi important que de connaître la loi. Ceux-ci s'efforcent depuis longtemps de réaliser des rapprochements et des comparaisons entre les formations de jugement, de manière plus ou moins empirique, afin de mieux conseiller leurs clients ayant affaire à tel ou tel juge ou formation de jugement.
44. Cette méthode était suffisante lorsqu'un avocat n'intervenait que devant un nombre limité de tribunaux, mais la levée progressive dans de nombreux pays des monopoles locaux du barreau, de même que la liberté d'installation au sein de l'Union européenne légitimement que tout avocat national et même européen puisse désirer connaître finement la jurisprudence de chaque juridiction nationale ou européenne, devant laquelle il est susceptible de plaider.

¹⁸ Voir CEDH, Z. c. *Finlandi* no. 22009/93, §§ 95 et suivants ainsi que le livre vert sur l'information émanant du secteur public dans la société de l'information de la Commission européenne COM(1995)585 (« *Si des précautions particulières ne sont pas prises, les bases de données jurisprudentielles, qui sont des instruments de documentation juridique, peuvent devenir des fichiers d'information sur les individus si ces bases de données sont consultées pour obtenir une liste des décisions de justice sur un individu spécifique plutôt que pour s'informer sur la jurisprudence* »)

¹⁹ Le profilage est défini à l'article 4 du RGPD. Il s'agit d'un traitement utilisant les données personnelles d'un individu en vue d'analyser et de prédire son comportement ou sa situation, comme par exemple déterminer ses performances au travail, sa situation financière, sa santé, ses préférences, ses habitudes de vie, etc.

45. Il n'est donc pas exclu, dans l'avenir, que des applications à forte valeur ajoutée, et donc forcément onéreuses, de *machine learning* puissent avoir une efficacité bien supérieure à l'expérience et au « bon sens » contentieux des avocats travaillant leurs dossiers de manière traditionnelle. L'utilisation de ces applications pourrait ainsi encore plus accentuer la distorsion de concurrence et une inégalité des armes entre des cabinets d'avocats ayant eu ou non recours de manière onéreuse à des tels logiciels d'analyse « prédictive » des jurisprudences.
46. Au prix d'un tel avantage compétitif, les risques sont réels de détournement du principe du juge naturel établi par la loi. En effet, les possibilités de « profilage du juge » dérivant du croisement des données publiques et privées pourraient permettre à des entreprises privées, et à leurs avocats, de se livrer à des pratiques encore plus accrues de *forum shopping*. Ce phénomène est déjà observé depuis longtemps aux Etats Unis, mais aussi en France pour le contentieux des délits de presse et d'atteintes à la vie privée par voie de presse, à l'occasion duquel le demandeur choisit déjà le tribunal qui lui paraît accorder les montants les plus élevés de dommages et d'intérêts.
47. Par ailleurs, un grand nombre de discours confondent l'*open data* avec les impératifs de publicité pesant sur un certain nombre d'informations publiques : il est ainsi parfois avancé que le nom des professionnels devraient obligatoirement figurer dans les données ouvertes au regard des impératifs de publicité et de transparence.
48. Or la mise à disposition de données informatiques de jurisprudence est totalement autonome des principes de publicité des originaux ou les copies certifiées des décisions. L'objectif de l'*open data* est de permettre un traitement automatisé de la jurisprudence au titre d'une licence d'exploitation non-onéreuse. Comme indiqué précédemment, cette mise à disposition s'effectue sous la forme d'une base de données informatique complète, peu claire et non compréhensible directement par le citoyen.
49. Cette mise à disposition ne permet évidemment pas de répondre à l'impératif de publicité du nom des professionnels ayant concouru à une décision spécifique. Rappelons qu'en droit, la mention du nom des juges dans les décisions de justice est une obligation commune aux Etats membres, en lien avec le principe de publicité de l'article 6§1 de la Convention européenne des droits de l'homme, qui vise à assurer de l'impartialité objective de juges (lesquels doivent être identifiables, régulièrement nommés et affectés aux fonctions qu'ils exercent) et du respect des règles de forme (publicité, collégialité par exemple)²⁰.
50. La réponse à la question de la légitimité de la publication du nom des professionnels²¹ en *open data* n'est donc pas à rechercher sur le terrain de l'obligation de la publication du nom des professionnels dans les décisions. Il semble plutôt que l'enjeu se situe dans la conciliation d'impératifs souvent opposés : d'une part, la transparence de l'action publique, en permettant aux citoyens de connaître et « d'évaluer » leurs juges, et, d'autre part, la protection de la vie privée des professionnels (dont la fonction ne saurait diminuer leurs garanties fondamentales dans ce domaine). Des défis se posent ainsi sur le terrain des garanties d'impartialité du juge, voire de l'institution judiciaire dans son ensemble, que les politiques d'*open data* viseraient à renforcer²² : comment préserver celles-ci *in concreto* des possibles tentatives de déstabilisation qui dériveraient du croisement des données nominatives des juges présentes dans la base de données avec d'autres sources (réseaux sociaux, sites marchands), et ceci dans le but de faire émerger d'hypothétiques biais politiques, religieux, etc ?
51. Ces questions ne se posent pas dans les mêmes termes partout en Europe et dépendent des spécificités du système judiciaire concerné (et de la nature de l'organe de gestion des carrières des juges), du caractère collégial ou non du jugement, et du niveau d'instance concerné. En Suisse par exemple où les juges sont élus, la publication est un gage de transparence et de

²⁰ CEDH, Vernes c. France, n. 30183/06 en ce qui concerne l'identification des juges ayant rendu la décision, Pretto c. Italie, n°. 7984/77, Grande Chambre, pour la publicité du jugement, Kontalexis c. Grèce, n. 59000/08, § 38, DMD GROUP, a.s., v. Slovakia, n°. 19334/03, § 66, Miracle Europe KFT v. Hungary, n° 57774/13, § 58 pour le droit à un tribunal établi par la loi en conjonction avec le droit à un juge impartial.

²¹ Ces questions peuvent aussi s'appliquer aux Rechtspfleger qui rendent des décisions judiciaires et aux greffiers, assistants du juge mentionnés dans la composition de la formation de jugement (quoique en moindre mesure).

²² Voir CEDH Previti c. Italie, n° 45291/06, §§ 249 et suivants qui rappelle les principes en matière d'impartialité objective du juge.

responsabilité sociale des juges vis-à-vis des citoyens ainsi que des groupes politiques. Ces informations sont d'ailleurs déjà disponibles dans les moteurs de recherche en ligne (qui ne sont pas à proprement parler de l'*open data*)²³.

52. Ces questions ne se posent également pas dans les mêmes termes selon les degrés de juridiction. L'intérêt de caractériser la jurisprudence de juges d'instances inférieures n'est peut être pas le même que pour des juges de Cours suprêmes ou de cours internationales. La Cour européenne des droits de l'homme autorise par exemple la recherche d'arrêts par le nom des juges de la formation de jugement, mais sans pour autant permettre le calcul de des statistiques concernant un juge en particulier²⁴. En revanche, dans les pays où les organes juridictionnels ne connaissent pas la pratique (existante dans cette juridiction internationale) de l'opinion dissidente, il peut paraître illégitime d'imputer à un juge, peut être mis en minorité lors du délibéré d'une formation collégiale, la responsabilité personnelle d'une telle décision.
53. Ces débats ont été bien circonscrits par une mission d'étude conduite en France par le professeur Loïc Cadiet. Les conclusions de la mission restent mesurées puisqu'elle ne recommande pas d'en prohiber le principe mais de réserver cette publication à certains types de contentieux et de l'exclure pour d'autres (pour les matières pénales spécialisées par exemple). La possibilité de publier seulement le nom des magistrats de la Cour suprême a été avancée, tout en soulignant la possibilité « d'aller sans retour » qu'une telle politique constituerait.
54. Un simple principe de précaution pourrait donc être appliqué en l'état afin d'évaluer, par type de contentieux et degré de juridiction, l'intérêt de la publication du nom des professionnels en base téléchargeable. De même, la possibilité d'exploiter cette information hors *open data*, par les institutions judiciaires même ou des tiers habilités, à des fins de connaissance de la jurisprudence n'est pas à omettre.
55. Une telle publication resterait en toute hypothèse à examiner au regard des règles européennes applicables en matière de protection des données personnelles, telles que celles de la Convention n°108 ou de l'Union européenne (RGPD et Directive 680/2016).

²³ Voir l'exemple du tribunal fédéral suisse, dont la jurisprudence est téléchargeable (<https://www.bger.ch/fr/index/juridiction/jurisdiction-inherit-template/jurisdiction-recht.htm>) ou des cantons (canton de Genève par exemple <http://ge.ch/justice/dans-la-jurisprudence>).

²⁴ Les décisions de la Cour européenne des droits de l'homme sont collégiales. Les publications comportent les éventuels avis dissidents.

3. Caractéristiques de fonctionnement de l'intelligence artificielle (apprentissage machine) appliquée aux décisions judiciaires

Le traitement du langage naturel et l'apprentissage machine sont les deux techniques au cœur du traitement des décisions judiciaires à l'aide de l'intelligence artificielle.

Dans la plupart des cas, l'objectif de ces systèmes n'est pas de reproduire un raisonnement juridique mais d'identifier les corrélations entre les différents paramètres d'une décision (par exemple dans un litige de séparation d'un couple : la durée de mariage, les revenus des époux, l'existence d'un adultère, le montant de la pension alimentaire prononcé, etc) et, en s'appuyant sur l'apprentissage machine, d'en déduire un ou plusieurs modèles, sur lesquels s'appuyer afin de « prédire » ou « prévoir » la future décision judiciaire.

Ce chapitre vise à apporter des éclaircissements sur la nature intrinsèque des logiciels qualifiés de « prédictifs » - parfois des versions « bêta », c'est-à-dire en phase de développement ultérieur - en évoquant tant leurs possibilités que leurs limites. Il visera à vulgariser les algorithmes dits d'apprentissage machine (*machine learning*) qui sont au cœur de l'analyse automatisée de la jurisprudence.

3.1 Les fonctionnalités théoriques d'un logiciel de justice dite prédictive

56. A titre d'introduction, il convient de revenir brièvement sur les fonctionnalités promises par ces logiciels dits prédictifs. Ils proposent d'établir des probabilités de succès (ou d'échec) d'une procédure devant un tribunal. Ces probabilités sont établies par la modélisation statistique des décisions antérieures via des méthodes provenant de deux domaines spécifiques de l'informatique : traitement de langage naturel (*natural language processing*) et apprentissage machine. Ces approches de modélisation statistique sont souvent qualifiées d'intelligence artificielle (IA) ; en réalité, il s'agit d'IA « faibles » (voire le glossaire)
57. Il doit être immédiatement précisé que le terme d'intelligence artificielle (IA) est débattu par les experts car il conduit à de nombreuses ambiguïtés. Le terme IA est désormais entré dans notre registre usuel de langage pour désigner un assemblage de sciences et de technologies assez diverses et qui permettent à des ordinateurs de battre les meilleurs champions du jeu de Go²⁵, de conduire une voiture, de converser avec les humains. Les chercheurs préfèrent pour leur part désigner les différentes applications par les technologies exactes qui les sous-tendent, dont l'apprentissage machine, et qualifie parfois l'ensemble de ces moyens d'IA très spécialisés comme des IA « faibles » (ou « modérées »). Ce qui les distingue d'un but ultime, encore totalement théorique, qui serait de créer une IA « forte », c'est-à-dire une machine auto-apprenante en capacité d'appréhender automatiquement le monde en général, dans l'ensemble de sa complexité.
58. Spécifiquement à la justice, les systèmes de justice dite prédictive ont pour ambition d'être utilisés par des directions juridiques, des assureurs (tant pour leurs besoins internes et que pour leurs assurés), ainsi que des avocats afin d'anticiper l'issue d'un contentieux. Ils auraient aussi vocation à assister les magistrats dans leurs prises de décisions.
59. Ils fournissent sous forme de représentation graphique des probabilités de succès pour l'issue d'un contentieux, en fonction de critères saisis par l'utilisateur (spécifiques pour chaque type de contentieux). Ces systèmes prétendent être en capacité de calculer les montants probables d'indemnisation par les tribunaux.

²⁵ <https://www.nature.com/articles/nature16961>

3.2 Le fonctionnement concret des intelligences artificielles : des machines statistiques construisant des modèles sur la base du passé

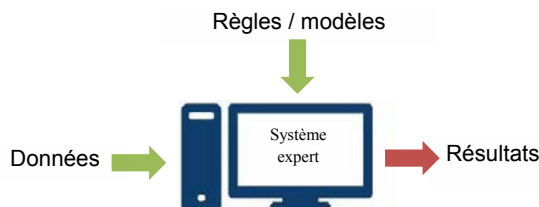
60. Une distinction paraît d'emblée devoir être opérée entre ce qui relève de la « *prédiction* » et ce qui relève de la « *prévision* ». La prédiction est l'acte d'annoncer (*prae*, avant - *dictare*, dire) par avance des événements futurs (par inspiration surnaturelle, par voyance ou prémonition). La prévision, quant à elle, résulte de l'observation (*visere*, voir) d'un ensemble de données afin d'envisager une situation future. Cet abus de langage et sa diffusion paraissent pouvoir être expliqués par un transfert du terme en provenance de l'univers des sciences « dures », où le terme fait référence à une variété de techniques de « *data science* » issues des mathématiques, des statistiques et de la théorie des jeux qui analysent des faits présents et passés pour réaliser des hypothèses sur le contenu d'événements futurs.
61. Il convient également de noter que la logique du fonctionnement d'un logiciel de justice prédictive repose essentiellement sur des méthodes génératives (communément appelées bayésiennes) ou discriminatoires qui tentent d'estimer la fourchette actuelle ou future des valeurs d'une variable (par exemple, le résultat d'un procès) à partir de l'analyse des exemples passés.
62. D'une manière générale, il convient également de garder à l'esprit « l'idée anthropomorphe que les machines calculatoires seraient intelligentes et que leurs concepteurs seraient parvenus à glisser un esprit à l'intérieur de leurs mécanismes »²⁶. Cette idée imprègne malheureusement encore de nombreuses analyses sur la justice dite prédictive qui prêtent à ces dispositifs des capacités immédiates ou futures de quasi réplique de l'intelligence humaine. Ce contexte, nourri chaque jour d'un lot de progrès révolutionnaire des IA²⁷, nous porte ainsi tous à aborder ces outils « prédictifs » avec une certaine part, consciente ou non, de mysticisme, en énonçant parfois que ce qui n'est pas tout à fait possible aujourd'hui le sera nécessairement demain.
63. En effet, les IA dites « fortes » de la littérature de science-fiction n'existent pas. Ce type d'IA, qui seraient douées non seulement d'intelligence mais aussi de conscience, demeurent purement fictionnel. Les systèmes d'apprentissage machine actuellement en cours de développement sont décrits comme des IA « faibles » et sont capables d'extraire des tendances complexes et d'apprendre à partir de grandes quantités de données de manière efficace et souvent avec des niveaux élevés de précision prédictive.
64. Pour dépasser toute considération instinctive ou fantasmée, ces technologies de traitement et d'analyse de l'information sont à appréhender autour de trois concepts :
- **l'IA n'est pas un objet unique et homogène** : il s'agit en réalité d'un assemblage de sciences et techniques (mathématiques, statistiques, informatique) en capacité de traiter des données pour concevoir des tâches très complexes de traitement informatique ;
 - **le moteur de l'IA ne produit pas de l'intelligence en soi mais fonctionne par une approche inductive** : l'idée est d'associer de manière plus ou moins automatisée un ensemble d'observations (entrées) à un ensemble de résultats possibles (sorties) à l'aide de diverses propriétés préconfigurées ; spécifiquement pour la « justice prédictive » le moteur construit des liens entre les différents groupes lexicaux composant les décisions judiciaires. Ces groupes sont corrélés entre ceux identifiés en entrée (les faits et la motivation) et ceux en sortie (le dispositif de la décision) et classifiés ;
 - **la fiabilité du modèle (ou fonction) construit dépend fortement de la qualité des données utilisées et du choix de la technique d'apprentissage machine**
65. Il convient de revenir brièvement sur les ancêtres de ces systèmes, les systèmes experts (SE), qui s'appuyaient pour leur part sur des règles de traitement écrites par un informaticien. Les SE ont connu un rapide développement à la fin des années 80 et dans les années 90, surtout dans

²⁶ Dominique Cardon, *A quoi rêvent les algorithmes, nos vies à l'heure des big data*, La République des idées, Editions du Seuil, 2015, p. 58.

²⁷ « AlphaZero : l'IA de DeepMind apprend et maîtrise le jeu d'échecs en moins de 4 heures », *Génération NT*, article publié le 8 décembre 2017, [En ligne], <https://www.generation-nt.com/alphazero-ia-deepmind-maitrise-echecs-4-heures-actualite-1948919.html> (page consultée le 14 décembre 2017).

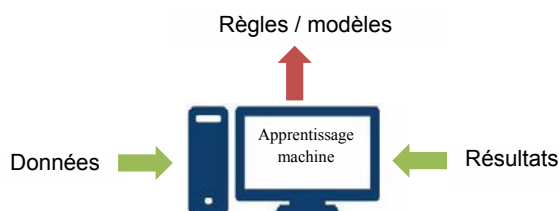
les domaines de la médecine et de la finance²⁸. Ces systèmes étaient capables de répondre à des questions spécialisées et effectuaient un raisonnement à partir de faits connus, en exécutant des règles d'encodage prédéfinies dans un « moteur ». Malgré le succès de SE comme *Deep Blue* face à Garry Kasparov lors d'une série de parties d'échecs en 1997, ces systèmes ont fini par échouer notamment parce qu'ils étaient dans l'incapacité d'interpréter « l'infinie variété des situations et des contextes »²⁹ et devenaient peu efficaces au-delà de 200 à 300 règles d'encodages, tant en terme de performance d'exécution que de maintenance (le raisonnement suivi par le système devenait quasiment impossible à appréhender pour ses concepteurs).

Graphique.2 : Les anciens systèmes experts étaient programmés avec des règles reproduisant la logique du raisonnement juridique



66. Aujourd'hui, l'idée n'est plus d'écrire des règles de raisonnement en miroir du raisonnement humain, comme pour les anciens systèmes experts³⁰, mais de laisser les systèmes d'apprentissage machine procéder eux-mêmes à l'identification de modèles statistiques existants dans les données et les mettre en correspondance avec des résultats spécifiques.

Graphique 3 : L'apprentissage machine produit seul des modèles en recherchant automatiquement des corrélations entre données et résultats.



67. Ce faisant, ces systèmes ne reproduisent pas et ne modélisent pas un raisonnement (par exemple juridique). A titre d'exemple les actuels traducteurs en ligne ne réalisent pas un raisonnement abstrait. Ils infèrent une estimation probable de la meilleure correspondance entre des groupes de lexicaux avec des traductions déjà réalisées. Les actions des utilisateurs

²⁸ Il peut être pris l'exemple du « High Frequency Trading (HFT) » ou transaction à haute fréquence, qui est un type transaction financière effectué à grande vitesse par des logiciels se basant sur des algorithmes. En 2013 et en Europe, 35 % des transactions étaient déjà réalisées avec le HFT. Aux USA, le HFT représente 70 % des volumes échangés sur le marché des actions. Au vu de l'automatisation des transactions, la chute est encore plus rapide en cas de krachs, comme ce fut le cas durant la crise financière de 2007.

²⁹ Hubert Dreyfus, *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, The MIT Press, 1992 cité par Dominique Cardon in « A quoi rêvent les algorithmes, nos vies à l'heure des Big Data », p. 59.

³⁰ Les règles de traitement écrites pouvaient s'appuyer sur du code spécifique, dans différents langages de programmation tels que le LISP ou sur des éditeurs permettant de modéliser des règles. Voir par exemple Radboud Winkels, « CLIME : Un projet de développement de serveurs juridiques intelligents », dans Danièle Bourcier, Patricia Hasset et Christophe Roquilly (dir.), *Droit et intelligence artificielle*, 2000, Romillat, p. 59.

contribuent évidemment à l'amélioration de la recherche de correspondance, mais l'algorithme d'apprentissage machine ne réalise pas à proprement parler une traduction en comprenant le sens des phrases traitées³¹.

68. Cet exemple montre la différence d'approche qui peut exister entre les juristes et les chercheurs : un juriste cherchera à comprendre l'effectivité au regard de règles préexistantes, ce que le chercheur ne pourra lui apporter qu'au travers du langage de sa science et de sa capacité à interpréter les milliers de paramètres issus d'un apprentissage machine.
69. En effet, l'apprentissage machine est un domaine des sciences informatiques dans lesquels les programmes des ordinateurs apprennent de l'expérience. Ces algorithmes font réaliser à une machine un processus d'entraînement, tel un enfant qui apprendrait dans son environnement. Sommairement, ces techniques d'apprentissage peuvent être supervisées ou non par un humain. La catégorie d'apprentissage machine la plus en vue est actuellement celle dite d'apprentissage non supervisé par renforcement : la machine récolte seule des « récompenses » virtuelles si l'action, dans un environnement donné, produit le résultat attendu. Les méthodes d'apprentissage machine qui incluent les réseaux de neurones (ou leur version plus complexe dite d'apprentissage profond - *deep learning*), ont été assez largement médiatisées au vu de leur autonomie et de leurs applications assez saisissantes comme celle parvenant à obtenir des scores élevés avec d'anciens jeux vidéo sous Atari 2600³² (sur la seule base du positionnement des pixels et des scores). D'autres applications concrètes de ces technologies affectent déjà notre quotidien et commencent à poindre dans le monde professionnel de la justice³³.
70. Toutefois, ces algorithmes restent très spécialisés pour une tâche et présentent des problèmes de discernement face à des situations chaotiques ou comportant un nombre insuffisant de données pour permettre une « prédiction » (comme la compréhension réelle du langage naturel). Dans le domaine des sciences sociales, auquel le droit et la justice appartiennent, l'échec apparaîtrait même inéluctable en absence de modèle convaincant de la cognition. Pour Pierre Lévy, « *l'intelligence artificielle se contente dans les faits de fournir une boîte à outils hétéroclite (règles logiques, syntaxes formelles, méthodes statistiques, simulations neuronales ou socio-biologiques...) qui n'offrent pas de solution générale au problème d'une modélisation mathématique de la cognition humaine* »³⁴. Ainsi, un réel apprentissage « prédictif » devrait en réalité s'appuyer sur une bonne représentation systémique du monde, qui se heurte, pour Yann LeCun, chercheur dans le domaine de l'intelligence artificielle, à une « *question fondamentale scientifique et mathématique, pas à une question de technologie* »³⁵.
71. Par ailleurs, l'originalité des systèmes actuels de traitement des *big data* n'est pas de prétendre reproduire notre modèle de cognition mais de réaliser une statistique des contextes sur une taille inédite de données, sans réelle garantie d'exclure les corrélations dites fallacieuses³⁶.
72. Alors qu'il est communément admis que les modèles statistiques et probabilistes sont améliorés par l'augmentation des données les alimentant, certains mathématiciens ont alerté sur les risques d'augmentation des corrélations dites fallacieuses, (c'est-à-dire des liens entre des facteurs n'ayant absolument aucun lien de causalité) dans le Big Data. Les mathématiciens Cristian Sorin Calude et Giuseppe Longo évoquent même le risque d'un « *déluge de corrélations fallacieuses dans le big data* » : « *Plus une base de données exploitée pour des corrélations est grande, plus grandes sont les chances de trouver des régularités récurrentes et plus grandes sont les chances*

³¹ Li Gong, « La traduction automatique statistique, comment ça marche ? », *Interstices.info*, publié le 29 octobre 2013, [En ligne], https://interstices.info/cms/nn_72253/la-traduction-automatique-statistique-comment-ca-marche (page consultée le 14 décembre 2017).

³² <https://www.nature.com/articles/nature14236>

³³ Intelligence artificielle d'IBM Watson équipant le domaine médical ou servant de moteur à « Ross », présenté comme un avocat « virtuel » – Ross, « Do more than humanly possible », [En ligne], <http://rossintelligence.com> (page consultée le 14 décembre 2017).

³⁴ Pierre Lévy, « Intelligence artificielle et sciences humaines », *Pierre Lévy's blog*, 8 octobre 2014, [En ligne], <https://pierrelevyblog.com/2014/10/08/intelligence-artificielle-et-sciences-humaines/> (page consultée le 30 décembre 2017).

³⁵ Yann LeCun, « Qu'est-ce que l'intelligence artificielle », recherche publiée sur le site du Collège de France, [En ligne], https://www.college-de-france.fr/media/yann-lecun/UPL4485925235409209505_Intelligence_Artificielle_Y._LeCun.pdf (page consultée le 14 décembre 2017).

³⁶ Dominique Cardon, *op. cit.*, p. 60.

de commettre des erreurs »³⁷. Ce qui peut apparaître en effet comme des régularités pour une IA (des liens récurrents entre différentes données, concepts, contextes ou groupes lexicaux) peut s'avérer en réalité aléatoire. Même si l'argument des deux mathématiciens ne doit pas être généralisé trop hâtivement, ceux-ci constatent que dans certains ensembles vastes de nombres, de points ou d'objets apparaissent des motifs réguliers aléatoires et il paraît impossible de les distinguer algorithmiquement des motifs révélant des causalités³⁸.

73. Enfin, les algorithmes, quelle que soit leur sophistication actuelle, demeurent nécessairement réduits aux interactions opérées dans la phase d'apprentissage et, partant, à leur passé. Le contenu et la qualité des flux de données entrant dans la composition des calculs est donc fondamentale pour avoir une compréhension des résultats obtenus et identifier des éventuels biais d'analyse. Les défis sont ici encore nombreux puisque, avec une approche exhaustive, l'analyse de la plus grande quantité possible de données relatives à une activité va produire des résultats dont le sens devra nécessairement être éclairé au regard de tous les facteurs ayant eu une influence. Avec une approche plus restrictive par échantillonnage de données entrantes, il existera aussi des risques au vu des biais propres aux arbitrages nécessaires pour sélectionner une donnée plutôt qu'une autre.

³⁷ Cristian Sorin Calude, Giuseppe Longo, « Le déluge des corrélations fallacieuses dans le big data », dans Bernard Stiegler (dir.), *La toile que nous voulons*, 2017, FYP éditions, p.156.

³⁸ Théorie dite de « Ramsey », Ronald L. Graham, Joel H. Spencer – « Ramsey Theory », *Scientific American*, vol.263, n°1, juillet 1990, p112-117 cité par Cristian Sorin Calude, Giuseppe Longo, « Le déluge des corrélations fallacieuses dans le big data », *ibid.*

4. Les intelligences artificielles peuvent-elles modéliser *a priori* un raisonnement juridique ?

Les intelligences artificielles contournent les difficultés rencontrées avec les anciens systèmes experts : elles ne tentent pas d'être un miroir d'un raisonnement juridique, dont la reproduction ne constitue pas pour elles un objectif. L'apprentissage machine conduit à effectuer des catégorisations entre les différents paramètres identifiés par les concepteurs ou ceux découverts par la machine.

74. Les systèmes juridiques continentaux sont très éloignés de l'idéal de rationalité que devait incarner par exemple le code civil de 1804 en France. Il existe une multitude de sources qui ne s'imbriquent pas de manière parfaitement cohérente et qui portent sur un ensemble de règles dont la signification demeure très indéterminée, ce que le théoricien du droit Herbert. L. A. Hart qualifiait de « texture ouverte du droit »³⁹. Les systèmes de Common Law, même réputés plus efficaces « économiquement » car s'adaptant dynamiquement aux nouveaux besoins juridiques⁴⁰, sont également évolutifs et n'offrent pas plus de sécurité juridique. Les règles juridiques ne connaissent donc pas une évolution linéaire, à l'inverse des lois empiriques, celles des sciences dites exactes, où chaque nouvelle règle vient en générale compléter les précédentes et non invalider un ensemble complet de raisonnement.
75. Rappelons que face à cette complexité générale des systèmes juridiques, les anciens systèmes experts informatiques ont rapidement trouvé leurs limites dès lors que l'on atteignait une imbrication de 200 à 300 règles logiques. Le découpage du droit sous forme de « règles de production » n'a pas suffi à donner une représentation valable de l'ensemble des connaissances et méthodes qui guident un juriste.
76. Si la modélisation informatique des systèmes juridiques positifs se révèle en conséquence impossible, il en va de même du raisonnement des juges. Comme l'a mis en lumière la théorie du droit, le raisonnement judiciaire est surtout affaire d'appréciation et d'interprétation, celle des faits prouvés et pertinents du litige, des règles de droit applicables (textuelle ou des précédents), dont la signification demeure, on l'a dit, très indéterminée⁴¹, mais aussi de l'interprétation subjective par les juges de la notion d'équité, qui devrait connaître en Europe de nouveaux développements avec l'exigence encouragée par la Cour européenne des droits de l'homme de « contrôle de proportionnalité ».
77. Xavier Linant de Bellefonds soulignait que, « la complexité du droit tient à son caractère téléologique et contentieux : deux raisonnements cohérents peuvent conduire à des arbitrages différents en fonction de deux ordres d'impératifs⁴². »
78. C'est que le fameux syllogisme judiciaire est davantage un mode de présentation du raisonnement juridique que sa traduction formelle. Celui-ci ne rend pas compte de l'intégralité du raisonnement tenu par le juge, lequel est en réalité ponctué d'une multitude d'éléments décisionnels, de choix, non formalisables *a priori* et parfois relevant de son pouvoir d'appréciation : quels sont les faits pertinents ? Ces faits sont-ils prouvés ? Quelle règle leur est applicable ? Quelle est le sens de cette règle au regard du cas à trancher ? Quelle source faire prévaloir entre une pluralité de sources en conflit ? La cohérence d'ensemble des décisions de justice n'est jamais assurée et relève davantage d'une mise en récit *a posteriori*, celle qu'opère notamment la doctrine juridique ou le juge dans sa motivation, ayant davantage pour fonction de convaincre du bien-fondé de telle ou telle solution que de décrire strictement et objectivement l'intégralité des étapes ayant concouru à la prise des décisions.

³⁹ Herbert L. A. Hart, *Le concept de droit*, Bruxelles, Facultés universitaires Saint-Louis, 1976.

⁴⁰ <https://www.contrepoints.org/2014/08/15/177160-common-law-contre-droit-civil-lexperience-francaise-de-lancien-regime>

⁴¹ Sur ces questions, Michel Troper, *La théorie du droit, le droit, l'Etat*, Paris, PUF, 2001, spéc. p. 69-84.

⁴² Xavier Linant de Bellefonds, « L'utilisation des systèmes experts en droit comparé », *Revue internationale de droit comparé*, 1994, Vol. 46, n° 2, p. 703-718.

79. Or ce travail d'interprétation est précisément ce que ne réalisent pas – et ne veulent pas réaliser – aujourd'hui les techniques d'apprentissage machine, puisqu'elles procèdent, nous l'avons vu, à un traitement automatisé où la corrélation de masses d'information est censée remplacer la compréhension des causalités réelles d'une décision. Celles-ci n'ambitionnent nullement de formaliser le raisonnement juridique mais elles espèrent que les modèles qu'elles capturent puissent anticiper les probables décisions d'un juge face dans des situations similaires.
80. Ces résultats auxquels les IA parviennent sont en réalité étrangers à la question de la conformité, en droit, de telle ou telle solution et ne peuvent discriminer entre des raisonnements conformes ou non en droit.
81. Un examen des travaux de l'University College of London (UCL) sur la jurisprudence de la Cour européenne des droits de l'homme confirme bien ce diagnostic. L'étude de l'UCL partait du postulat qu'un simple modèle d'apprentissage automatique pouvait prédire l'issue d'une affaire avec une précision de 79 % pour ce tribunal particulier. Le modèle issu de l'apprentissage machine s'est avérée plus efficace sur la partie descriptive des faits des décisions étudiées que sur la partie de motivation relative à l'application de la Convention au cas d'espèce⁴³.
82. L'examen des termes corrélés au constat d'une « violation » (avec une pondération positive) et ceux corrélés à une non-violation de la Convention européenne des droits de l'Homme (avec une pondération négative) éclaire sans ambiguïté la mécanique à l'œuvre, qui n'est en rien comparable avec un raisonnement juridique.

Graphique 4 : Illustration des travaux de l'UCL – Poids théorique alloué à des mots ou termes en fonction de leur lien avec des constats de violation ou de non violation de la Convention européenne des droits de l'Homme.



83. Car, précisément, le choix des faits pertinents et leur qualification constituent l'un des éléments de la décision du juge. Autrement dit, l'étude de UCL n'a été en réalité capable de produire une probabilité qu'avec un matériel lexical en grande partie issu du raisonnement et de la motivation du juge et non pas avec celui du requérant, qu'elle a assemblée en fonction des seules fréquences d'apparition. Leur IA a donc établi une forte probabilité de correspondance entre des groupes de mots et une décision d'ores-et-déjà formalisée. Elle ne saurait en aucun cas reproduire seule le raisonnement des juges européens ni surtout prédire une issue, sur la base par exemple d'un

⁴³ Travaux sur un échantillon de 584 décisions de la Cour européenne des droits de l'Homme : Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiu-Pietro, Vasileios Lamos, « Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective », publié le 24 octobre 2016, [En ligne], <https://peerj.com/articles/cs-93/> (page consultée le 14 décembre 2017).

récit brut d'un futur requérant devant la Cour de Strasbourg, dont la requête sera soumise à un examen de recevabilité très strict (près de 70 356 requêtes ont été déclarées irrecevables ou rayées du rôle en 2017⁴⁴) reposant en grande partie sur l'application de standards d'appréciation (importance et sérieux du grief, etc.) ménageant un espace décisionnel important.

⁴⁴ Source : Analyse statistique 2017 de la CEDH, janvier 2018.

5. Les IA peuvent-elles expliquer *a posteriori* le comportement des juges ?

L'explication *a posteriori* d'un comportement d'un juge, notamment la révélation de biais nécessiterait d'identifier tous les facteurs potentiellement causatifs au travers d'un cadre interprétatif et de procéder à un travail d'analyse contextualisé. Le fait, statistiquement, que la garde des enfants est plus souvent confiée aux mères qu'aux pères ne démontre pas nécessairement un parti pris des juges mais conduit à devoir mobiliser différentes disciplines issues des sciences sociales pour éclairer ce phénomène.

84. D'un point de vue scientifique, expliquer un phénomène ou, pour ce qui nous concerne, un comportement humain, revient à déterminer les mécanismes de causalité ayant conduit à ce comportement en présence d'un certain nombre de données contextuelles.
85. Cela nécessite, de manière très schématique, la constitution préalable d'un cadre interprétatif, lui-même né de l'observation répétée de la réalisation de ce type d'événements ou de comportements en présence de certains facteurs ou éléments. Ce cadre interprétatif est constitué par les hypothèses ou points de vue adoptées par les différentes disciplines des sciences sociales. Il s'agit d'une étape supplémentaire d'analyse dont peuvent être nourris les algorithmes, mais à laquelle ils ne procèdent pas seuls.
86. Certaines *legaltechs* sont allées relativement loin et ont pensé pouvoir identifier d'éventuels biais personnels des magistrats et nourrir des suspicions de partialité. La mise en *open data* du nom de certains présidents de formation de jugement des tribunaux administratifs et cours administratives d'appel en France a ainsi permis de construire un indicateur du taux de rejet de recours contre les obligations de quitter le territoire français prises par les autorités administratives. Certains commentateurs avaient alors soutenu avec beaucoup d'enthousiasme que la prétendue impartialité des juges était ainsi « mise à mal par l'intelligence artificielle ».
87. Mais peut-on réellement parvenir à de telles interprétations sur la base d'un traitement algorithmique des décisions de justice ? Caractériser un biais personnel dans la prise de décision d'un magistrat (en dehors naturellement des cas d'expression personnelle et publique de ce magistrat dans l'affaire concernée) consiste à fournir une explication d'un comportement, en l'occurrence d'une décision, par les traits de personnalité, les opinions ou encore la confession d'un juge. Or, on l'a dit, une telle explication causale ne peut être purement et simplement déduite du résultat probabiliste fourni par les algorithmes. Cela requiert au contraire un travail d'analyse supplémentaire devant permettre d'isoler, parmi les nombreux facteurs corrélés (dont l'identité de membres de la formation de jugement), ceux véritablement causatifs. Par exemple, le fait qu'un juge aux affaires familiales fixe statistiquement davantage la résidence habituelle d'un enfant chez la mère ne traduit pas nécessairement un parti-pris de ce magistrat en faveur des femmes mais davantage l'existence de facteurs psycho-sociaux, économiques voire culturels propres au ressort de la juridiction, tels que le temps de travail de chacun des parents, le montant de leurs revenus, l'offre locale de garde collective, la scolarisation ou non de l'enfant, la refondation ou non par un des parents d'un nouveau couple ou même tout simplement l'absence d'intérêt d'un des deux parents à s'occuper d'un jeune enfant.
88. De même, on ne peut pas comparer purement et simplement les décisions rendues en matière d'expulsion du territoire par une juridiction administrative située à proximité d'un grand centre de rétention avec celles d'une juridiction qui connaît ce contentieux de manière marginale.
89. D'ailleurs quel que soit le lieu de la juridiction, la question de la « jurisprudence » d'un juge unique de permanence qui ne traite qu'occasionnellement d'un type de contentieux, mais qui s'inspire (ou non) de la jurisprudence de ces collègues est particulièrement intéressante et elle peut poser légitimement la question de l'égalité des citoyens devant la production judiciaire, mais elle relève plus d'une culture judiciaire de la « collégialité » à réinventer ou à préserver que d'un « fichage » ou d'une « stigmatisation » via les outils de *machine learning*.

90. Que déduire, enfin, de la personnalité du président d'une formation de jugement collégiale, alors que son nom est la seule information nominative visible dans les décisions ouvertes de la justice administrative ?
91. Comment comparer également deux approches philosophiques et culturelles distinctes de la décision de justice, lorsque dans certains pays européens ou en France en matière de justice administrative règnent la culture du « précédent » et une connaissance fine par les juges des bases de données factuelles de toutes les décisions de 1^{ère} et de 2nde instance (base Ariane), alors que dans d'autres pays ou systèmes sont privilégiés l'indépendance « intellectuelle » de chaque juridiction, de même que la volonté de traiter au « cas par cas » chaque situation ?
92. Fournir une véritable explication d'une décision de justice nécessite donc d'analyser beaucoup plus finement les données contingentes de chaque affaire et les règles de droit applicables, plutôt que de nourrir le vain espoir que de la masse des liens nait le sens.

6. Quelles applications pour la justice civile, commerciale et administrative ?

L'état de développement des techniques d'apprentissage machine ne permet pas à ce jour de parvenir à des résultats fiables s'agissant de la « prédiction » des décisions judiciaires. En revanche, leur application dans le champ de la justice civile, commerciale et administrative est à considérer pour la constitution de barèmes ou la résolution précontentieuse de litiges en ligne, dès lors qu'un recours ultérieur au juge demeure possible.

93. On le voit, la question que pose en premier lieu une telle utilisation des intelligences artificielles ne porte pas tant sur le fait de savoir si elles sont bénéfiques ou néfastes, souhaitables ou non, mais sur celui de savoir si les algorithmes que l'on nous propose peuvent parvenir ou non au type de résultat recherché. Les conclusions de l'expérimentation conduite dans les cours d'appel de Douai et de Rennes en France démontrent bien qu'en présence d'un discours séduisant de promotion d'un produit présenté comme né de l'IA peut cacher des défauts rédhibitoires de conception et des résultats d'analyse totalement erronés.

94. Indépendamment de la qualité du logiciel testé, l'anticipation de la décision du juge apparaîtrait cependant, en matière civile, commerciale et administrative, comme une opportunité, pour des préoccupations parfois très différentes, tant pour les responsables de politiques publiques judiciaires que pour les professionnels privés du droit.

95. Quelle que soit la tradition juridique du pays, l'aléa judiciaire, c'est-à-dire le risque de voir sa demande en justice satisfaite ou rejetée, suscite l'envie de pouvoir être quantifié à l'aide de ces nouvelles applications technologiques.

96. Les avocats (ou les services juridiques d'une entreprise) y voient la possibilité d'apporter à leurs clients un conseil mieux éclairé en évaluant de manière empirique et systématique les chances de succès d'une procédure, mais aussi de favoriser la conclusion de transactions permettant, le cas échéant, d'échapper à un long et coûteux procès. Certains assureurs proposent déjà l'utilisation de systèmes prédictifs à leurs clients pour évaluer les mérites de leur affaire⁴⁵.

97. Dans le même temps, les décideurs publics voient l'opportunité de mieux réguler les flux de nouvelles procédures dans les tribunaux et de se doter d'un levier de réduction des coûts de fonctionnement de la justice. Il s'agirait ainsi d'inciter les justiciables à recourir par ce biais à des modes alternatifs de résolution de litiges (conciliation, médiation ou encore arbitrage)⁴⁶.

Les expérimentations menées en France

A l'initiative du Ministère de la justice, les deux cours d'appel de Rennes et de Douai ont accepté de tester sur divers contentieux d'appel au printemps 2017 une version d'un logiciel présenté comme de « prédictibilité de la justice » et en réalité d'analyse des décisions en matière civile, sociale et commerciale de l'ensemble des cours d'appel françaises.

Alors que ces données internes et exhaustives de jurisprudence leur étaient déjà accessibles gratuitement depuis de nombreuses années (base JURICA), et qu'elles avaient été mises spécialement à la disposition de la société éditrice par le Ministère, les deux cours s'étaient vues proposer d'expertiser l'intérêt d'une analyse chiffrée, présentée comme novatrice, des sommes allouées par les juridictions, ainsi que d'un classement géographique des écarts constatés pour des requêtes et des procès similaires.

L'ambition affichée du logiciel était donc la création d'un outil d'aide à la décision, aux fins de réduire si nécessaire une trop grande variabilité des décisions des tribunaux, et ce au nom du principe d'égalité des citoyens devant la loi. Le retour d'expérimentation, contradictoirement débattu entre les deux cours d'appel, le Ministère de la justice et la *legaltech* à l'origine du produit a malheureusement objectivé l'absence de plus-value de la version testée du logiciel pour le travail de réflexion et de prise de décision des magistrats utilisateurs.

Plus grave, des biais de raisonnement du logiciel ont été révélés qui aboutissaient à des résultats aberrants ou inappropriés, en raison d'une confusion entre de simples occurrences lexicales des motivations judiciaires et les causalités qui avaient été déterminantes dans le raisonnement étudié des juges.

⁴⁵ Le logiciel Predictice est proposé aux assurés d'Allianz par exemple.

⁴⁶ Voir à ce titre en France le rapport d'information n° 495 (2016-2017) au nom de la commission des lois du Sénat, déposé le 4 avril 2017 du sénateur Philippe BAS.

98. L'approche déjà existante dans de nombreux systèmes judiciaires, d'une harmonisation des décisions par des barèmes dans de nombreuses matières (divorce, licenciement, réparation du préjudice corporels) semble pouvoir être renouvelée par une approche probabiliste ou actuarielle⁴⁷. Des services alternatifs de résolution de litiges en ligne ont même vu le jour, entre autres pour aider à l'évaluation du montant de réparation pour des petits litiges. Ces approches intéressantes ne sont toutefois pas dénuées de biais et elles ne doivent pas priver le citoyen d'accéder à un juge ni remettre en cause le principe du contradictoire.

6.1 Une nouvelle puissance de calcul pour la constitution de barèmes

99. Les **modalités de calcul de barèmes** dans diverses matières civiles (indemnisation du préjudice corporel, prestation compensatoire, indemnité de licenciement par exemple) paraissent pouvoir être considérablement enrichies, en combinaison avec d'autres techniques de traitement et sous réserve de nombreuses précautions de conception et d'utilisation (effet dit « performatif »)⁴⁸.
100. En effet, rappelons ce que Jean-Paul Jean, président du groupe de travail évaluation de la CEPEJ qualifiait de « défi qualitatif » lors d'un colloque relatif à l'*open data* tenu en 2016 en France : le traitement opéré par la *machine learning* ou tout autre modalité de traitement devrait pouvoir se réaliser sur la base d'originaux certifiés, dont l'intégrité aura été vérifiée, et qui auront été enrichis pour distinguer l'important de l'insignifiant⁴⁹.
101. L'autre risque auquel ont déjà été confrontés les concepteurs de ces outils est celui de se livrer à du *data-snooping*, c'est-à-dire à ne sélectionner en amont que des données de manière à ce qu'elles puissent être significatives au regard de grilles d'analyse prédéterminées, par exemple en éludant de l'échantillon des décisions se prêtant mal à des corrélations de séquences linguistiques par la *machine learning* ou tout autre méthode (des décisions sans exposé du litige ou peu motivées par exemple).
102. Mais si l'on calcule un barème, ne fait-on pas déjà un peu de « prédiction » ? La frontière peut sembler relativement ténue sauf si l'on distingue clairement la finalité de la démarche : l'objectif ici n'est pas de délivrer une information prescriptive mais d'informer d'un état de fait.
103. Au prix de ces précautions méthodologies et d'utilisation, les barèmes apparaissent pour le traitement de certains contentieux un puissant outil d'harmonisation de la jurisprudence. Auparavant calculés sur des échantillons plus ou moins restreint de décisions, les outils issus de l'IA rendent possible l'exploration d'une quantité plus importante de décisions et sont susceptibles de permettre la constitution d'échelles plus précises, selon un rythme de mise à jour plus important.

6.2 Les résolutions de litiges en ligne

104. Les tribunaux européens font tous face, selon des degrés différents, à des contentieux civils répétitifs de faible valeur. L'idée de faciliter leur traitement par les technologies de l'information et/ou de les externaliser des tribunaux est assez largement partagée. La Grande-Bretagne, les Pays-Bas, la Lettonie représentent autant d'exemples de pays qui ont déjà mis en œuvre ou sont sur le point de mettre en œuvre ce type des solutions plus ou moins automatisées⁵⁰. Pour les

⁴⁷ S'agissant de l'approche actuarielle, l'offre de CaseLaw Analytics annonce réaliser plus une évaluation de risque que la prédiction d'une solution à un litige.

⁴⁸ L'effet performatif ou d'autoréalisation est le risque pour un système de produire progressivement en sortie les mêmes résultats en influant sur les producteurs d'information en entrée ; cet effet est souvent évoqué en ce qui concerne les barèmes en matière judiciaire qui, renseignés de décisions basées sur ces échelles, tendraient à n'être représentatifs que d'eux-mêmes.

⁴⁹ J.-P. Jean, Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence, Colloque du 14 octobre 2016 tenu à la Cour de cassation, <https://www.courdecassation.fr/IMG//Open%20data,%20par%20Jean-Paul%20Jean.pdf>, page consultée le 17 mars 2018.

⁵⁰ Voir à ce titre le système de résolution de litiges en ligne en Grande-Bretagne - <https://www.judiciary.gov.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version1.pdf>. Voir également le système "AP" aux Pays-Bas, qui émet des décisions automatisées basées sur des permis précédemment accordés, et qui a donné lieu à des litiges au niveau national et devant la CJCE : deux affaires (C-293/17 et C-294/17) sont portées devant le Conseil d'Etat des Pays-Bas (agriculteurs / conservation de la nature contre les Pays-Bas) pour déterminer si un système (Programme Régulant l'azote) est autorisé ou non à décider

créances transfrontières, l'Union européenne a mis en place, par le règlement n° 524/2013, un cadre commun disponible sur internet (« European small claims »).

105. Le périmètre de ces services de résolution de litiges en ligne (*online dispute resolution – ODR*) semble toutefois avoir progressivement évolué : de services en ligne restreints à des mesures alternatives de règlement de litiges (*alternative dispute resolution*) pour un traitement avant la saisine de la juridiction, ces offres s'introduisent de plus en plus dans le processus juridictionnel lui-même au point d'offrir des services de « tribunal électronique ».⁵¹ Elles ne concernent pas uniquement des litiges de faible valeur, mais s'appliquent également au contentieux fiscal ou à celui relatif aux services de sécurité sociale, ou encore aux procédures de divorce.
106. Pour les promoteurs de ces solutions, qui intéressent un certain nombre de professions judiciaires et le secteur privé, l'accès à la justice pourrait être significativement amélioré par une solution large combinant ODR et IA (ou à tout le moins des systèmes experts, voir *supra* chapitre 3 pour la distinction). L'idée est d'accompagner les plaignants dans un diagnostic automatisé du litige au travers d'un certain nombre de questions traitées par la machine, aboutissant à des propositions de solution. Les travaux du laboratoire de Cyberjustice de Montréal qui intègre les différentes phases précontentieuses et contentieuses dans un processus informatique pour des conflits de faible intensité (cour des petites créances du Québec par exemple) illustrent bien cette hybridation⁵². Les résultats en termes d'efficacité et de qualité paraissent selon les concepteurs exceptionnels.
107. Mais sur quelle base serait calculée l'éventuelle indemnisation proposée par un tel système ? Quelle méthode ? L'algorithme traite-t-il l'information de manière loyale ? Les propositions ont-elles vocation à pouvoir être discutées contradictoirement avec l'aide d'un tiers formé et certifié ? Un accès au juge est-il toujours possible ? Certains auteurs voient même dans la généralisation de ces modes de règlement de litiges une nouvelle manifestation du « solutionnisme » numérique, c'est-à-dire du recours systématique aux technologies pour tenter de résoudre des problèmes n'en relevant pas nécessairement⁵³. Rappelons aussi qu'un Europe, un cadre réglementaire contraignant pour les États membres et plus protecteur a été récemment mis en place : le RGPD, dans son article 22, prévoit explicitement la possibilité pour la personne concernée de refuser de faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, sauf exceptions⁵⁴.

si les agriculteurs et autres personnes violent la directive habitat. Récemment, la Cour de justice de l'Union européenne à Luxembourg a répondu aux demandes de décision préjudicielle concernant ces affaires (ECLI:EU:C:2018:882).

⁵¹ Darin Thompson, *Creating new pathways to justice using simple artificial intelligence and online dispute resolution*, Osgoode Hall Law School of York University

⁵² <http://www.cyberjustice.ca/projets/odr-plateforme-daide-au-reglement-en-ligne-de-litiges/>

⁵³ Evgeny Morozov, *Pour tout résoudre cliquez ici*, éditions FYP, cité par David Larrousserie, *Contre le « solutionnisme » numérique*, Le Monde, 6 octobre 2014, https://www.lemonde.fr/sciences/article/2014/10/06/contre-le-solutionnisme-numerique_4501225_1650684.html

⁵⁴ Article 22, 1 du Règlement UE 2016/679 : « La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé » ; des exceptions sont prévues (tels que le consentement de la personne concernée) mais des « mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée » doivent être mises en œuvre par le responsable du traitement dont « droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision ».

Voir, dans le même sens, la Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, telle qu'elle sera amendée par le Protocole adopté en mai 2018 quand ce dernier entrera en vigueur. L'article 9§1, sous a) prévoit le principe que « Toute personne a le droit de ne pas être soumise à une décision l'affectant de manière significative, qui serait prise uniquement sur le fondement d'un traitement automatisé de données, sans que son point de vue soit pris en compte. Par dérogation à ce principe d'interdiction, l'article 9§2 précise que « le paragraphe 1.a ne s'applique pas si la décision est autorisée par une loi à laquelle est soumis le responsable du traitement et qui prévoit également des mesures appropriées pour la sauvegarde des droits, des libertés et des intérêts légitimes de la personne concernée ».

Le rapport explicatif précise (§75) : « Il est essentiel que toute personne susceptible d'être soumise à une décision purement automatisée ait le droit de contester cette décision en faisant valoir de manière effective son point de vue et ses arguments. En particulier, la personne concernée doit avoir la possibilité de prouver l'inexactitude éventuelle des données à caractère personnel avant leur utilisation, l'inadéquation du profil qu'il est prévu d'appliquer à sa situation particulière ou d'autres facteurs qui auront un impact sur le résultat de la décision automatisée. Tel est notamment le cas lorsque l'application d'un raisonnement algorithmique, en conduisant à la limitation d'un droit, au refus d'un avantage social ou à l'évaluation de leur capacité d'emprunt sur le seul

108. Il faut donc arriver à bien mesurer, au cas par cas, les potentiels bénéfiques d'un système d'ODR, son degré d'intégration dans un processus juridictionnel complet (du précontentieux au véritable contentieux) ainsi que le rôle plus ou moins décisif de l'IA dans l'exécution du processus.
109. En effet, pour les offres d'ODR existant déjà en amont des processus juridictionnels, leur place paraît relativement claire : il s'agit, hors du tribunal, de contribuer à la mise en œuvre de conciliation, médiation, d'arbitrage. Ces services peuvent également intervenir dans le cadre d'une procédure contentieuse sous le contrôle du juge avant qu'il ne tranche le litige sur le fond (une telle phase est parfois rendue obligatoire pour certains litiges).
110. D'autre part, l'apport effectif de l'IA est à mesurer. S'agit-il seulement d'utiliser de l'apprentissage machine pour établir des barèmes indicatifs ou prescrire une solution ? S'agit-il réellement d'IA ou de système expert ou du simple enchaînement de règles logiques ? En toute hypothèse, il conviendrait de pouvoir assortir ces systèmes d'impératifs de transparence, de neutralité et de loyauté⁵⁵.
111. Enfin, il faut également examiner la manière dont les plaignants sont incités à utiliser le système : existe-t-il une confusion ou non dans la dénomination même de l'offre ? Si l'on parle de tribunal, il doit s'agir de la forme d'organisation définie par la Convention européenne des droits de l'homme et non une simple justice privée aux simples apparences d'une justice étatique⁵⁶. Le recours au juge est-il clairement rendu possible ? Aux Pays-Bas des contrats privés d'assurance maladie semblaient prévoir par défaut un recours à un ODR avant toute saisine de la justice.
112. Le groupe de travail médiation de la CEPEJ (CEPEJ-GT-MED) a initié en 2018 des premières réflexions sur les apports des technologies de l'information pour la résolution de litiges avec des méthodes alternatives. Le CDCJ conduit actuellement des travaux approfondis sur les ODR afin d'identifier les potentialités de ces outils mais également les points problématiques au regard d'éventuelles violations des articles 6, 8 et 13 de la Convention européenne des droits de l'homme.

6.3 Les principales garanties à réaffirmer dans le cadre de la procédure civile, commerciale et administrative

Droit d'accès à un tribunal

113. La mise à disposition d'outils de résolution de litiges en ligne ne devrait pas remettre en cause le droit d'accès à un tribunal au sens de l'article 6, même si ce droit n'est pas absolu et se prête à des limitations implicites⁵⁷. En matière civile par exemple, chaque justiciable possède le droit de soumettre à un tribunal toute contestation relative à ses « droits et obligations de caractère civil »

fondement du logiciel, a pour effet de stigmatiser des individus. La personne concernée ne pourra néanmoins pas exercer ce droit si la décision automatisée est prévue par la loi à laquelle le responsable du traitement est soumis, qui prévoit des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée ».

Voir également l'article 9§1, sous c), de la Convention 108 modernisée, qui prévoit le droit de la personne concernée « d'obtenir, à sa demande, connaissance du raisonnement qui sous-tend le traitement de données, lorsque les résultats de ce traitement lui sont appliqués ». Le rapport explicatif de la Convention modernisée (§77) précise : « Les personnes concernées ont le droit d'obtenir connaissance du raisonnement qui sous-tend le traitement des données, y compris les conséquences de ce raisonnement et les conclusions qui peuvent en avoir été tirées, en particulier lors de l'utilisation d'algorithmes pour une prise de décision automatisée, notamment dans le cadre du profilage. Par exemple, dans le cas d'un système d'évaluation de leur solvabilité par notation, les emprunteurs ont le droit d'obtenir connaissance de la logique sur laquelle repose le traitement de leurs données et qui aboutit à la décision d'octroi ou de refus du crédit, au lieu d'être simplement informés de la décision elle-même. La compréhension de ces éléments contribue à l'exercice effectif d'autres garanties essentielles comme le droit d'opposition et le droit de recours auprès de l'autorité compétente ». On retrouve également dans la RGPD ce « droit d'obtenir des informations utiles concernant la logique sous-jacente » (article 13§1, sous f ; article 14§2, sous g ; article 15§1, sous h).

⁵⁵ Charlotte Pavillon, Préoccupations au sujet d'un juge numérique, nrc.nl, <https://www.nrc.nl/nieuws/2018/01/19/zorgen-om-populaire-digitaal-rechter-a-1588963>

⁵⁶ Scarlett-May Ferrié, Les algorithmes à l'épreuve du procès équitable, document consulté sur <http://lexis360.fr>, téléchargé le 09/07/2018, § 27-38

⁵⁷ Art.6 §1 « 1. Toute personne a droit à ce que sa cause soit entendue (...) par un tribunal (...) qui décidera (...) du bien-fondé de toute accusation en matière pénale dirigée contre elle. (...) » ; pour les limitations voir Deweer c. Belgique, § 49 ; Kart c. Turquie [GC], § 67.

⁵⁸. L'Assemblée parlementaire du Conseil de l'Europe a adopté en 2015 une résolution relative à « L'accès à la justice et internet: potentiel et défis » dans laquelle elle appelle « à veiller à ce que les parties qui engagent une procédure de règlement de litige en ligne conservent un droit d'accès à une procédure d'appel devant les tribunaux qui satisfasse à l'exigence d'un procès équitable énoncée à l'article 6 de la Convention »⁵⁹.

Principe du contradictoire

114. S'agissant des barèmes, il semble impératif de rendre accessible aux citoyens et, en premier lieu, aux parties à un procès un certain nombre d'informations quantitatives (nombre de décisions traitées pour obtenir le barème par exemple) et qualitatives (origine des décisions, représentativité des échantillons sélectionnés, distribution des décisions entre différents critères comme le contexte économique et social) afin de saisir le sens des échelles construites, d'en mesurer les éventuelles limites et de pouvoir en débattre devant un juge.

Egalité des armes

115. L'utilisation de moyens technologiques ne devrait pas conduire à aggraver le déséquilibre entre les parties. L'utilisation de moyens numériques pourrait en effet faciliter les recours pour certains opérateurs (institutionnels, entreprises disposant de moyens, personnes alphabétisées numériquement) et, au contraire, poser des difficultés à certains types de population plus précarisées ou moins éduquées sur le numérique. Il est important que tout individu ne soit pas laissé seul derrière son écran, mais qu'il puisse être renseigné de la possibilité d'avoir accès à un conseil juridique et d'être accompagné dans sa démarche le cas échéant.

Impartialité et indépendance des juges

116. On peut se demander si « la norme issue du nombre » évoquée précédemment à la section 2.2. ne produirait pas d'effets indirects sur l'indépendance et l'impartialité des magistrats, et notamment dans des systèmes dans lesquels l'indépendance du système judiciaire n'est pas totalement acquise. Dans ces systèmes, le risque que cette « norme issue du nombre » puisse avoir un effet indirect de pression sur les magistrats lors de la prise de décision et favoriser une uniformisation des magistrats, voire un contrôle par l'exécutif de ceux qui dérogent à la norme ne peut pas être écarté.

Droit à l'assistance d'un avocat

Nous avons évoqué au début de ce chapitre les avantages dérivant de l'application des outils de justice dite prédictive pour les avocats, et notamment la possibilité d'apporter à leurs clients un conseil mieux éclairé en évaluant de manière empirique et systématique les chances de succès d'une procédure. Imaginons toutefois un cas où les chances de succès pour le justiciable soient extrêmement réduites : est-ce que cela pourrait avoir un effet sur la décision de l'avocat d'assister son client ? La pratique professionnelle devrait viser à minimiser le risque que des personnes nécessitant un conseil juridique puissent au final, en être privées.

⁵⁸ CEDH, *Golder c. Royaume-Uni*, §§ 28-36

⁵⁹ Résolution 2054 (2015) de l'Assemblée parlementaire du Conseil de l'Europe (APCE), 10 novembre 2015, <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-fr.asp?fileid=22245&lang=fr>

7. Les enjeux spécifiques à la justice pénale : prévention des infractions, du risque de récidive et évaluation de la dangerosité

Même sans être conçus pour être spécifiquement discriminatoires, le recours aux statistiques et aux IA en matière pénale a démontré un risque de résurgence de doctrines déterministes au détriment des doctrines d'individualisation de la peine, largement acquises depuis 1945 dans la plupart des systèmes judiciaires européens.

117. L'utilisation des sciences et technologies d'IA en matière pénale pose des difficultés spécifiques car leur application pourrait traduire certains débats publics actuels sur le caractère prétendument prédictible des comportements infractionnels. Ce débat avait pourtant paru tranché durablement durant une trentaine d'années dans un certain nombre de pays européens. En Italie par exemple, l'article 220, paragraphe 2, du Code de procédure pénale exclut expressément le recours à une expertise aux fins d'établir « *l'habitude ou le professionnalisme dans le crime, la tendance à commettre un crime*, le caractère et la personnalité de l'accusé et, en général, les qualités psychologiques de l'accusé, indépendamment des causes pathologiques ». En France, la doctrine dite de « *défense sociale nouvelle* » développée par Marc Ancel a été fondatrice du droit pénal : à une approche simplement punitive et déterministe s'est substitué un système de réadaptation sociale postulant prévenir la commission d'infraction en évitant que soient réunies les conditions de la délinquance. Cette approche est partagée par un nombre d'instruments européens sur les politiques criminelles qui mettent l'accent sur les objectifs de rééducation et de réinsertion des sanctions pénales⁶⁰.
118. Les outils en matière de justice pénale seraient donc à concevoir dans le respect de ces principes fondamentaux de réhabilitation⁶¹ et en rappelant le rôle du juge en matière d'individualisation de la peine, sur la base d'éléments objectifs de personnalité (formation, emploi, prise en charge médico-sociale régulière) sans autre forme d'analyse que celle opérée par des professionnels spécifiquement formés, comme les agents de probation par exemple. Les techniques de *Big Data Analytics* pourraient être utilisées par ces professionnels pour centraliser et collecter des informations sur la personne accusée d'un crime ou délit qui pourraient se trouver auprès d'institutions et organismes différents et qui auraient ensuite besoin d'être examinées par un juge, parfois dans des délais très courts (par exemple dans le cadre des procédures accélérées de jugement).

7.1 Les outils utilisés par les services d'enquête avant le procès pénal

119. Les instruments qualifiés de « *police prédictive* » (avant judiciarisation ou avant la saisine d'un tribunal) connaissent déjà un essor rapide et commencent à être connus par le grand public (pensons par exemple aux « *no fly lists* » qui sont en réalité une application de « *big data analytics* » rassemblant et analysant des données sur de potentiels terroristes dans le but de prévenir la commission d'actes ou encore aux algorithmes employés pour détecter des faits de fraude ou de blanchiment).
120. D'une manière générale, de nombreux outils informatiques sont couramment utilisés pour prévenir la commission d'actes criminels (en identifiant les lieux de prédilection des passages à l'acte ou leurs auteurs potentiels) ou pour les poursuivre de manière plus efficace⁶². On retrouve dans la première catégorie des instruments qualifiés de « *police prédictive* » qui servent à empêcher certains types d'infractions ayant des éléments de régularité dans leur occurrence comme le cambriolage, la violence dans la rue, le vol des/dans les véhicules. Cette qualification est liée à la capacité de ces technologies de déterminer précisément où et quand ces infractions pourraient être commises, et de reproduire ces informations sur une carte géographique, sous forme de « *points chauds* » qui sont visualisés en temps réel par les patrouilles de police. On parle d'ailleurs de « *cartographie criminelle prédictive* ». La plupart des logiciels utilisés dans ce

⁶⁰ See European Court of Human Rights, Grand Chamber, *Winter vs United Kingdom*, paras 114 – 118.

⁶¹ L'utilisation de l'IA à des fins de traitement et de réhabilitation (pour collecter par exemple des données sur les traitements administrés ou les méthodes de réintégration en prison) paraît en revanche devoir être encouragée.

⁶² Voir à ce propos Ales Zavrsnik, "Big Data, crime and social control", page 194 et suivantes, qui énumère de manière détaillée une série d'instruments utilisés par les services de police en Europe et aux États Unis.

domaine se base sur des éléments de localisation historique des crimes, comme des rapports de police, mais des nouvelles technologies encore plus puissantes combinant des données variées et provenant de différentes sources sont également testées⁶³. Ces instruments, qui affichent des taux d'efficacité très persuasifs, auraient par ricochet des effets dissuasifs en ce qui concerne la commission d'infractions dans les zones entourant les « points chauds » et positifs pour les politiques publiques⁶⁴.

121. Il faut toutefois relativiser les capacités de « prédiction » de ces outils, qui montrent leurs limites à l'égard de crimes de nature moins régulière ou visant des lieux différents, comme le terrorisme. Par ailleurs, une de leurs faiblesses est l'effet de « cercle vicieux » et de « prophéties auto-réalisatrices » : les quartiers considérés à risque attirent davantage l'attention de la police et la police y détecte davantage de crimes, ce qui mène à une surveillance excessive par la police de ces communautés qui les habitent⁶⁵. Enfin, des interrogations quant à une possible « tyrannie de l'algorithme » qui pourrait minimiser, voire progressivement remplacer le jugement humain ne sont pas totalement absentes au sein des services de police eux-mêmes, même si pour l'instant, la technologie est présentée comme restant au service de l'humain pour mieux l'outiller dans la prise de décision⁶⁶.
122. En outre, les « Big Data Analytics » sont de plus en plus appliqués dans le cadre de la poursuite du crime. Des outils comme Connect, qui est utilisé par la police britannique et qui analyse des milliards de données produites dans le cadre des transactions financières pour retrouver des corrélations ou des schémas d'opération, ou encore la base de données internationale sur l'exploitation sexuelle des enfants (ICSE DB), gérée par Interpol, qui aide à identifier des victimes et / ou des auteurs par l'analyse, par exemple, des meubles et autres objets présents dans des images abusives, ou l'analyse des bruits de fond dans des vidéos, se sont montrés particulièrement efficaces dans le combat contre le crime. S'agissant de Connect par exemple, des recherches nécessitant autrefois des mois d'enquête peuvent désormais être effectuées en quelques minutes, avec un niveau de complexité et sur un volume de données très importants.
123. La doctrine s'interroge néanmoins sur la logique managériale de la réponse au crime apportée par ces outils de prédiction, dans laquelle « une analyse approfondie des raisons du crime devient moins important que le « faire quelque chose ici et maintenant » et dans laquelle, dans un contexte de réduction des budgets disponibles et dans une spirale vicieuse, la police doit assurer le même niveau de protection du public, mais avec un personnel, un équipement et des ressources limités⁶⁷.

7.2 Les outils durant le procès pénal

124. L'utilisation d'outils « prédictifs »⁶⁸ par le juge dans le cadre d'un procès pénal s'avère très peu répandue en Europe.

⁶³ Par exemple, dans le cadre du projet « E-Security – ICT for knowledge-based and predictive urban security » (<http://www.esecurity.trento.it/>), qui a été mené dans la ville de Trento en Italie de novembre 2012 à mai 2015, une base de donnée réunissant des informations relatives aux crimes dénoncés à la police, aux résultats d'enquêtes menés par la mairie portant sur la victimisation et la sécurité réelle et perçue par les citoyens, des éléments sur le désordre urbain physique et social provenant de la police, ainsi que d'autres variables relatives à la « Smart City (par exemple des informations relatives aux conditions sociodémographiques, le décor urbain, l'illumination nocturne, la présence de caméras de surveillance, de transports publics etc) » a été créée pour mieux outiller le travail de prévention du crime et d'amélioration de la sécurité urbaine. Les responsables du projet ont témoigné de la fiabilité des techniques utilisés, qui permettraient de prédire la commission d'actes criminels avec un taux de succès d'environ 60-65 % et qui aiderait à mieux combattre le crime dans un contexte général de pénurie de ressources. Par ailleurs, de tests menés au Royaume Uni dans le cadre d'un projet pilote visant à prévoir les lieux de possibles cambriolages, vols, et assaut montrent que les projections du logiciel utilisé, dénommé PREDPOL, étaient précises dans 78 % des cas, contre le 51 % des techniques traditionnelles.

⁶⁴ L'indication d'une concentration géographique du crime aiderait les forces de police à mieux s'interroger sur les facteurs environnementaux de prédisposition de celui-ci dans la zone identifiée (par exemple l'éclairage, la présence de commerces etc) et à prévoir des réponses adéquates en concertation avec d'autres acteurs.

⁶⁵ Predicting crime, LAPD style, The Guardian, 25 June 2014.

⁶⁶ How technology is allowing police to predict where and when crime will happen, The Independent, 7 October 2017.

⁶⁷ Aleš Završnik, "Big Data, crime and social control", page 196.

⁶⁸ Dans la doctrine, ces outils sont souvent qualifiés de « justice algorithmique » ou « justice automatisée », ou encore « justice simulée ».

125. HART – Harm Assessment Risk Tool, a été mis au point avec l'Université de Cambridge et est à présent testé au Royaume Uni. Cette technologie fondée sur l'apprentissage automatique a été entraînée à partir de cinq années d'archives de la police de Durham, comprises entre 2008 et 2012. En apprenant des décisions prises par les policiers pendant cette période, et de la récidive ou non de certains suspects, la machine est censée être capable d'évaluer le risque – faible, moyen ou élevé – des suspects, sur la base notamment d'une trentaine de facteurs, dont certains non liés au crime commis (par exemple le code postal, le genre).
126. Dans les tests initialement menés en 2013, dans lesquels le comportement des suspects a été observé sur une période de deux ans après la commission du crime, les prévisions de HART se sont révélées efficaces à hauteur de 98 % en cas de risque faible et de 88 % en cas de risque élevé de récidive. Dans cette phase d'expérimentation, HART aura une valeur purement consultative pour le juge; par ailleurs, des audits du fonctionnement de HART et de la fiabilité de ses conclusions seront régulièrement menés par la police.
127. Même s'il s'agit du seul outil « prédictif » recensé en Europe à ce jour, il est opportun de s'interroger sur les défis auxquels les décideurs publics pourraient se trouver confrontés dans un futur proche si ce type d'application était testé à une plus large échelle, au regard notamment de l'expérience américaine.
128. Aux Etats-Unis⁶⁹, l'ONG ProPublica a révélé les effets discriminatoires de l'algorithme utilisé dans le logiciel COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), dont l'ambition est d'évaluer le risque de récidive lorsque le juge doit déterminer la peine d'un individu.
129. Cet algorithme, qui a été développé par une entreprise privée et qui doit être utilisé obligatoirement par le juge dans certains états fédérés américains, regroupe 137 questions auxquelles répond la personne concernée, complétées par son cahier judiciaire. Les questions sont assez variées et incluent la présence d'un téléphone à la maison, la difficulté de paiement de factures, les antécédents familiaux, l'histoire criminelle du prévenu⁷⁰. L'algorithme note la personne sur une échelle de 1 (faible risque) à 10 (haut risque). Il représente une aide à la prise de décision judiciaire, ses conclusions n'étant qu'une des variables à considérer par le juge lors de la définition de la peine.
130. Or, les populations afro-américaines se sont vues attribuer par l'algorithme – sans que cet effet ne soit naturellement recherché par les concepteurs⁷¹ - un taux élevé de risque de récidive deux fois supérieur à celui des autres populations dans les deux ans qui suivaient l'application de la peine; inversement l'algorithme considèrerait que d'autres populations semblaient beaucoup moins susceptibles de réitérer une infraction⁷². Il doit être évidemment rappelé que ce type d'interprétation, fallacieuse, ne révèle en réalité que la fragilité sociale et économique de certains

⁶⁹ Une recherche de 2015 avait recensé une soixantaine d'outils prédictifs aux Etats Unis.

⁷⁰ Il y a d'autres algorithmes qui ont été élaborés sur la base d'observations critiques exprimés par la doctrine (voir le chapitre suivant) qui se basent sur des variables plus restreintes, plus directement liées au crime commis et moins à la race, le genre ou la condition socioéconomique. Citons à titre d'exemple le Public Safety Assessment Tool utilisé en 30 juridictions américaines.

⁷¹ Cet effet purement discriminatoire est en réalité explicable par le « calibrage » relativement permissif du modèle de l'algorithme qui créé beaucoup de « faux positifs ».

⁷² Les populations noires étaient plus fréquemment classées à haut risque, alors qu'elles ne récidivaient pas dans les deux ans suivant la condamnation ; les populations blanches à l'inverse, qui étaient plus fréquemment classées à faible risque, commettaient des infractions dans les deux ans suivants. En bref, l'algorithme surestimait le risque de récidive pour les noirs et le sous-estimait pour les blancs (les « faux positifs » étaient majoritairement des noirs, alors que les « faux négatifs » étaient en majorité des blancs). En réponse aux allégations de ProPublica, NorthPointe (devenue Equivant depuis cette polémique) a répondu que les populations blanches et noires étaient également représentées lorsqu'on considèrerait les « vrais positifs », c'est-à-dire les personnes ayant effectivement récidivés. **La question de comment concilier à la fois la précision de l'algorithme dans la détection de la récidive et la nécessité d'éviter des effets discriminatoires vis-à-vis des populations noires a été la source d'un intense débat dans la doctrine** ; voir en particulier Chouldechova A (2016), « Une juste prédiction avec un impact disparate: une étude sur le biais dans les instruments de prédiction de la récidive », disponible sur : <http://arxiv.org/abs/1610.07524>; et aussi « Bias in criminal risks scores is mathematically inevitable, Researchers say », disponible sur <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>. **Ce débat traduit par ailleurs la question de la légitimité d'une société privée, en dehors de tout contrôle institutionnel, d'arbitrer entre deux exigences opposées : celle de défense de la société, d'une part, et celle du respect des droits des individus, d'autre part.**

groupes de populations qui ne sont évidemment pas criminogènes par nature. Des chercheurs du Dartmouth College ont par ailleurs démontré que ce type d'algorithmes ne produisait pas de valeur ajoutée puisque des humains sans aucune expérience judiciaire arrivaient à reproduire exactement la même évaluation en interprétant simplement le questionnaire.

131. Par ailleurs, l'opacité des processus de fonctionnement des algorithmes par les entreprises privées (qui revendiquent leur propriété intellectuelle) a été une autre source d'inquiétude. Si l'on tient compte du fait qu'elles tiennent leurs données sources des autorités étatiques elles-mêmes, leur absence d'esprit de responsabilité et de quant aux règles relatives à la protection des données (accountability) vis-à-vis des citoyens pose un problème démocratique majeur. Les chroniques ont montré que le public est informé des opérations relatives aux big data de manière accidentelle, sporadiquement, lorsqu'il y a des fuites ou des erreurs : exemplaire à cet effet est le cas de ProPublica qui a révélé les failles dans l'algorithme de COMPAS suite au refus de l'entreprise propriétaire de partager celui-ci, et qui a dû saisir à la fois les autorités publiques pour avoir accès aux données et engager son propre scientifique pour examiner l'algorithme.

7.3 Les défis du « prédictif » en matière pénale

132. Nous avons vu dans les sections précédentes que l'état de développement des outils prédictifs en Europe dans le domaine pénal est très varié. Si les instruments qualifiés de « police prédictive » connaissent un essor rapide et commencent à être connus même par le grand public, le constat n'est pas le même s'agissant de l'application de ce type d'outils par le juge dans le cadre du procès pénal. S'agissant des instruments à disposition des services de poursuite, des considérations sur leurs avantages et inconvénients ont déjà été exprimées précédemment. Etudions maintenant les outils spécifiques au procès pénal.
133. Tout d'abord, il est important de bannir une perspective unique d'efficacité ou, au contraire, d'inefficacité de ces outils. Les exemples illustrés précédemment montrent qu'il peut y avoir des opportunités formidables mais aussi de véritables risques dans l'application des nouvelles technologies, faute des précautions nécessaires. Les décideurs publics et les acteurs judiciaires doivent être particulièrement vigilants et jouer un rôle actif dans le développement de ces technologies : un regard continu est nécessaire pour déterminer la réelle efficacité, efficacité et éviter des conséquences imprévisibles. Cette exigence est encore plus forte en matière pénale en raison de l'incidence directe sur la liberté personnelle de l'individu⁷³.
134. Cela dit, les possibilités offertes mais aussi les inconvénients de l'application de tels outils dans le domaine judiciaire devraient être attentivement mesurés.
135. Leurs promoteurs font souvent valoir leur caractère neutre et le fait qu'ils se baseraient sur des moyens factuels et objectifs qui aideraient à rendre la justice plus précise et transparente. Un autre grand atout serait leur efficacité, parfois dépassant les capacités humaines, qui ne peut qu'être extrêmement précieuse dans un contexte général de réduction de fonds publics, voire de pénurie de ressources.
136. Leurs détracteurs expriment en revanche des préoccupations sur l'impact et le danger potentiel que ces outils auraient sur une justice individualisée et égale pour tous : l'inclusion de variables algorithmiques comme l'histoire criminelle et les antécédents familiaux a comme conséquence que le comportement passé d'un certain groupe peut décider du sort d'un individu, qui est, bien entendu, un être humain unique, avec un contexte social, une éducation, des compétences spécifiques, un degré de culpabilité et des motivations spécifiques qui le mènent à commettre un crime⁷⁴. Par ailleurs, ils font valoir que la décision humaine peut se fonder sur des valeurs et des considérations (par exemple sociétales) qui ne seraient pas retenues par la machine. Par exemple, un juge pourrait décider d'ordonner la liberté sous caution d'une femme délinquante tout

⁷³ Un extrait de la décision de la Cour Suprême de Winsconsin dans l'affaire *Winsconsin vs Loomis* peut offrir inspiration aussi au niveau européen : « Il est important de considérer que des outils tels que COMPAS continuent de changer et d'évoluer. Les préoccupations que nous abordons aujourd'hui pourraient très bien être atténuées à l'avenir. Il incombe au système de justice pénale de reconnaître que dans les mois et les années à venir, des données de recherche supplémentaires seront disponibles. Des outils différents et meilleurs peuvent être développés. À mesure que les données changent, notre utilisation d'outils basés sur des preuves devra également changer. Le système de justice doit suivre la recherche et évaluer continuellement l'utilisation de ces outils. »

⁷⁴ Aleš Završnik, "Big Data, crime and social control", page 196.

en ayant bien présent le risque de récidive, sur la base d'une hiérarchisation des valeurs, par exemple en donnant une plus grande importance à son rôle de mère de famille et de garde des enfants, alors que l'algorithme arriverait à déterminer de manière plus précise le risque de réitération de l'infraction mais n'arriverait pas à opérer une telle hiérarchisation d'impératifs.

137. En matière pénale, les risques de discrimination sont aussi possibles lorsqu'on considère que ces outils, qui sont construits et interprétés par l'humain, peuvent reproduire des inégalités injustifiées et déjà existantes dans le système de justice pénale concerné : au lieu de corriger certaines politiques problématiques, la technologie finirait par les légitimer. Comme évoqué précédemment, l'ONG ProPublica⁷⁵ a révélé les effets discriminatoires de l'algorithme utilisé dans COMPAS⁷⁶, qui prévoyait un risque de récidive des populations noires deux fois supérieur à celui des populations blanches dans les deux ans qui suivaient l'application de la peine, et inversement considérait ces dernières beaucoup moins susceptibles de réitérer l'infraction. Certains auteurs mettent toutefois en relief les potentialités d'utilisation des algorithmes pour corriger les inégalités du système pénal concerné : puisque ils reflètent l'humain, il pourraient comme un miroir révéler les erreurs inhérentes à la prise de décision et faciliter leur correction par les acteurs⁷⁷.
138. A la lumière des considérations exprimées ci-dessus, lorsqu'on utilise l'algorithme dans le cadre d'un procès pénal il paraît essentiel de garantir pleinement le respect du principe d'égalité des armes et de présomption d'innocence établis par l'Article 6 de la CEDH: la partie concernée devrait avoir accès et pouvoir remettre en cause la validité scientifique, le poids donné aux différents éléments et les éventuelles conclusions erronées de celui-ci, et cela chaque fois que l'algorithme est proposé à un juge ou utilisé par lui avant qu'il ne prenne sa décision.
139. A cet égard, on remarque une différence entre l'Europe et les Etats Unis en ce qui concerne le droit d'accès à l'algorithme : si aux Etats Unis les autorités judiciaires sont encore réticentes à reconnaître pleinement ce droit et opèrent une mise en balance entre les intérêts privés (et notamment la protection de la propriété intellectuelle) et les droits de la défense, en Europe, le cadre est plus protecteur du fait du RGPD, qui établit un droit d'information sur la logique sous-jacente des décisions prises sur la base d'algorithmes⁷⁸.
140. Les considérations exprimées précédemment quant aux effets potentiellement négatifs de ces outils sur l'impartialité du magistrat sont également valables en matière pénale : un juge qui décide contre la prédiction d'un algorithme prend vraisemblablement des risques d'assumer une responsabilité accrue. Il ne semble pas irréaliste d'imaginer que les juges seraient récalcitrants à prendre cette charge supplémentaire, notamment dans les systèmes dans lesquels leurs mandats ne sont pas permanents mais assujettis au vote populaire⁷⁹, ou dans lesquels leur responsabilité personnelle (disciplinaire, civile voire pénale) est susceptible d'être recherchée, surtout si leurs garanties statutaires en matière disciplinaire sont insuffisantes.

⁷⁵ www.propublica.org/article/technical-response-to-northpointe

⁷⁶ D'autres algorithmes, se focalisent sur d'autres éléments plus directement liés à l'infraction commise.

⁷⁷ Mojca M. Plesnicar and Katja Sugman Stubbs, « Subjectivity, algorithms and the courtroom ».

⁷⁸ Article 15, 1. h) du Règlement UE 2016/679 : « La personne concernée a le droit d'obtenir du responsable du traitement « ... » les informations suivantes : « ... » l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée ».

⁷⁹ Mojca M. Plesnicar and Katja Sugman Stubbs, « Subjectivity, algorithms and the courtroom ».

8. Questions spécifiques relatives à la protection des données personnelles

L'utilisation des algorithmes pose la question de la protection des données personnelles traitées. Un simple principe de précaution devrait conduire à mener des politiques d'évaluation des risques.

141. Afin de permettre de saisir tout le potentiel offert par l'utilisation des algorithmes tout en veillant au respect des principes de protection des données, il conviendrait d'appliquer le principe de précaution et de mettre en place des politiques préventives concernant les risques potentiels liés à l'utilisation des données traitées par ces algorithmes et à l'impact de leur utilisation sur les personnes, et à plus large échelle sur la société.
142. Le principe de légitimité du traitement des données personnelles et l'obligation de prévenir ou de minimiser l'impact du traitement des données sur les droits et les libertés fondamentales des personnes concernées devraient imposer la réalisation préalable d'une évaluation des risques de l'impact potentiel du traitement des données par intelligence artificielle sur les droits et libertés fondamentales des personnes. Une telle évaluation devrait permettre de prévoir la mise en œuvre de mesures appropriées, notamment dès la conception (by-design) et par défaut (by default) destinées à atténuer les risques identifiés.
143. Rappelant que les données à caractère personnel doivent être traitées pour des finalités déterminées et légitimes, elles ne doivent pas être utilisées de manière incompatible avec ces finalités et elles ne doivent pas non-plus faire l'objet d'un traitement ultérieur que la personne concernée puisse considérer comme étant inattendu, inapproprié ou contestable (principe de loyauté). La problématique de la réutilisation des données personnelles qui auraient été rendues largement accessibles impose donc la plus grande prudence.
144. La conception du traitement de données effectué par les algorithmes devrait permettre de minimiser la présence de données redondantes ou marginales et d'éviter tout biais caché potentiel et tout risque de discrimination ou d'impact négatif sur les droits et libertés fondamentales des personnes concernées.
145. Dans cet environnement d'intelligence artificielle, les droits des personnes concernées revêtent une importance singulière et le contrôle que chacun de nous doit pouvoir garder sur ses informations personnelles implique que les droits suivants puissent être exercés : le droit de ne pas être soumis à une décision automatisée affectant de manière significative la personne concernée sans que son point de vue soit pris en compte, le droit d'obtenir connaissance du raisonnement qui sous-tend le traitement de données réalisé par les algorithmes, le droit de s'opposer à un tel traitement et le droit de disposer d'un recours.

9. Les potentialités et les limités des outils de justice dite prédictive

Le terme « justice prédictive » est à réfuter car il est ambigu et fallacieux. Ces outils relèvent de méthodes d'analyse de la jurisprudence, en s'appuyant sur des méthodes statistiques qui ne restituent en rien un raisonnement juridique mais peuvent tenter de le décrire. Les biais de l'analyse, s'ils ne peuvent être totalement supprimés, sont à identifier. Le processus de conception et d'utilisation doit s'intégrer dans un cadre éthique parfaitement clair.

146. Il a déjà été souligné, au chapitre 3, le caractère ambigu et fallacieux du concept de « justice prédictive » et comment celle-ci opère un lent glissement dans l'esprit collectif, en nous portant à croire que des machines, dénuées d'affect, seront un jour mieux à même de fiabiliser l'acte de juger. Les promesses de celle-ci ont besoin plus que jamais d'être examinées d'une manière objective et scientifique, en s'appuyant sur des solides bases de recherche fondamentale, pour en détecter d'éventuels limites. Rappelons à ce titre que les risques d'interprétations faussées du sens des décisions de justice s'avèrent extrêmement élevés sur la base des seuls calculs statistiques opérés ; ce constat est aggravé par l'absence de compréhension précise des liens opérés entre les données et le présence évidente de corrélations fallacieuses, non discernables dans des masses importantes de données.
147. Par ailleurs, la soi-disant neutralité des algorithmes est un mythe, leurs créateurs transférant, de manière consciente ou non, leurs propres systèmes de valeurs. Ainsi Eric Sadin, philosophe, rappelait que « les systèmes algorithmiques, sous leur allure efficace et impersonnelle, matérialisent de façon imperceptible les intentions de leurs concepteurs ou commanditaires, induisant un pouvoir agissant et asymétrique de personnes sur la vie d'autres personnes ». En ce sens, Aurélien Grosdidier, chercheur, estime que « un algorithme, en lui-même, n'est capable de rien d'autre que de nous permettre — au mieux — de saisir une partie de l'intention de son concepteur » et élargit le questionnement à l'ensemble de la chaîne de traitement de l'information (intention du concepteur, réalisation du code informatique, exécution du code informatique et contexte d'exécution puis maintenance). Ce constat est par ailleurs partagé par Aleš Završnik, criminologue, qui souligne comme les étapes de construction et d'interprétation des algorithmes sont effectuées par l'homme et pour l'homme et ne peuvent pas échapper aux erreurs, aux préjugés, aux valeurs et aux intérêts humains, et à une représentation humaine du monde, quelle que soit la manière dont ils sont conçus.
148. En dépit de ces fortes limites, doit-on pour autant négliger l'apport d'une technologie à la puissance inégalée ? Les mathématiciens Cr. S. Calude et G. Longo soulignent eux-mêmes dans leur étude relative au big data que « la portée limitative ou négative de [leurs] résultats, comme cela arrive souvent, ne détruit pas la science des données mais ouvre la voie pour de plus grandes réflexions », dont « le défi d'une nouvelle méthode scientifique, plus étendue, capable d'intégrer à la fois les nouveaux instruments algorithmiques et les outils classiques » en accompagnant les traitements d'une « évaluation rigoureuse des preuves ». Comme souligné en introduction, l'utilisation des IA est de nature à apporter une aide extrêmement significative aux professionnels, dont les magistrats et les avocats, mais également au grand public notamment si elles permettent un jour de forger des outils de recherche et d'analyse documentaire sans équivalent en matière législative, réglementaire, jurisprudentielle et doctrinale et de créer des liens dynamiques entre toutes ces sources . Mais ce type d'application sort du champ d'étude de cet article puisqu'il ne s'agit alors en aucun cas de « prédire » l'issue d'un contentieux mais d'analyser une jurisprudence dans un temps et un espace donné.
149. Nous avons également envisagé au chapitre 6, sous réserve d'une représentativité d'échantillons sélectionnés et traités, la construction de barèmes beaucoup plus précis de la moyenne ou la médiane des sommes allouées, mutatis mutandis, dans divers domaines (pensions alimentaires, prestations compensatoires, réparation du préjudice corporel, indemnités de licenciement, etc.). Ces barèmes qui relèvent plus d'une démarche de type « conférence de consensus » que d'une analyse moyenne de l'existant apportent d'ores-et-déjà une aide non négligeable à la décision et au conseil sans pouvoir se substituer à la loi elle-même. Comme évoqué précédemment, le risque est que, faute de représenter statistiquement la réalité ni de pouvoir prédire quoi que ce soit, les résultats des logiciels de « justice prédictive » ne soient érigés en normes en dehors de toute validation par le système juridique et en contradiction avec celui-ci.

150. Evoquons enfin une proposition qui consisterait à inclure dans les systèmes « prédictifs » la possibilité de pouvoir « dézoomer » à l'envi. Moins que d'enfermer l'utilisateur dans une probabilité (ou des groupes de probabilité), l'idée serait de lui permettre de naviguer dans les corrélations ayant conduit le système à proposer cette évaluation et de pouvoir s'en extraire par la sélection d'autres concepts ou groupes de mots plus pertinents ou d'exclure des corrélations fantaisistes. Pour reprendre l'exemple de l'UCL, cela consisterait à proposer une représentation graphique des différents termes retenus par le système (avec leurs poids respectifs) pour conclure à une violation (ou non-violation) et d'autoriser l'emprunt d'autres chemins en proposant la sélection d'autres termes ou groupes lexicaux.
151. Pour audacieuse et séduisante que soit cette proposition, elle suppose que les professionnels eux-mêmes (juges, avocats, universités) s'en emparent collectivement pour en éprouver la faisabilité et qu'ils ne laissent pas des opérateurs privés concevoir seuls, avec quelques scientifiques non contrôlés, des logiciels et des modes de raisonnement ou de calculs abscons ou verrouillés.
152. Les promesses ambitieuses (et non réalisées) de certaines legaltechs ne doivent donc pas occulter les immenses potentialités des technologies et le besoin d'applications adaptées et construites en lien direct avec la recherche scientifique et universitaire ainsi qu'avec l'ensemble des professionnels du droit, tels les magistrats, greffiers, avocats, notaires, huissiers et experts du terrain.
153. Dans ce contexte dynamique, il paraît essentiel, premièrement, de ne pas prendre des décisions hâtives et de débattre préalablement des enjeux et des applications concrètes qui pourraient être faites de ces instruments dans les systèmes judiciaires, aussi bien que de les tester dans une première phase. Par ailleurs, une institution judiciaire adaptée à son temps deviendrait une institution en capacité de fonder, d'administrer et de garantir à la fois une véritable **cyberéthique** tant au secteur public que privé de même que d'imposer une totale transparence et loyauté de fonctionnement des algorithmes susceptibles de participer un jour à la prise de décisions judiciaires.

10. La nécessité d'un débat public approfondi autour de ces outils antérieurement à la mise en place des politiques publiques de développement de ceux-ci. L'impérieuse nécessité d'une cyberéthique pour encadrer le développement des algorithmes d'intelligence artificielle dans le respect des droits fondamentaux

Les enjeux de l'intégration de ces outils dans un processus décisionnel judiciaire justifient de vulgariser auprès de l'ensemble des publics concernés les concepts en cause. Il convient de mettre en place un cadre éthique favorable à un rapide développement de l'IA qui intègre, dans ses processus même de conception, des mécanismes prévenant les éventuels biais et discriminations.

10.1 L'importance de débattre, tester et continuellement revoir l'application de ces outils antérieurement à la mise en place de politiques publiques

154. Les enjeux liés à la mise en place d'outils de justice prédictive sont tellement nombreux et à multiples facettes qu'ils exigent une approche pondérée par les décideurs publics.
155. Tout d'abord, il paraît essentiel d'organiser un débat public autour de ces questions, qui réunisse à la fois les concepteurs des outils et les professionnels de la justice. Les conseils de la magistrature, les associations professionnelles des magistrats, les barreaux pourront sans doute y contribuer et faciliter l'identification des potentialités et des aspects plus controversés. Par ailleurs, les écoles des magistrats et des avocats peuvent jouer un rôle clé de sensibilisation des professionnels de la justice autour de ces questions, afin qu'ils puissent mieux comprendre et contribuer utilement aux évolutions en cours.
156. Entreprendre une recherche fondamentale sur les applications proposées et les tester – pour comprendre leurs potentialités et leurs faiblesses, mais aussi pour pouvoir ultérieurement les développer et les adapter aux besoins – semble aussi essentiel. Un droit de regard sur les éléments et les caractéristiques des instruments proposés par le privé (ou ceux développés par des instituts publics indépendants et spécialisés, une solution qui devrait être encouragée) semble tout aussi indispensable afin que le service public de la justice puisse assurer efficacement sa mission. Une évaluation rigoureuse des résultats du test devrait être menée avant un déploiement à plus large échelle et une intégration dans les politiques publiques. Une évaluation régulière de l'impact de ces outils sur le travail des professionnels de la justice semble aussi hautement recommandable.

10.2 La mise en place d'un cadre éthique

157. A titre liminaire, n'envisager qu'un encadrement législatif ou réglementaire des finalités de traitement des IA n'est pas suffisant dans un contexte numérique, par essence totalement transfrontalier. Une attention scrupuleuse à la nature des données en *open data* et à leur qualité est en revanche de nature à minimiser les risques de croisements inadéquats et à renforcer la pertinence des résultats des traitements automatisés. Spécifiquement en ce qui concerne le nom des professionnels, un principe de précaution simple consisterait à prohiber leur diffusion publique dans des bases brutes structurées au regard des risques de détournements de finalités. Il ne s'agirait pas de limiter l'accès à l'information déjà traitée (restitution unitaire de la composition de la formation de jugement par exemple) mais de filtrer les données brutes librement mises à disposition. Distinguer en somme l'accès à l'information de l'accès à la base de données, manipulable à l'envie.
158. Les chercheurs Buttarelli et Marr ont souligné comment les « big data » ont besoin d'un grand contrôle et d'une grande protection. D'autres chercheurs (Pasquale, Morozov) ont mis l'accent sur la nécessité de mettre en place de procédures transparentes de déploiement des big data, et de manière plus générale des IA dans le domaine judiciaire, car les solutions proposées ne peuvent jamais représenter la vie dans sa complexité.

159. Le développement de règles de cyber-éthique pour guider l'activité des opérateurs dans le secteur et pour faire valoir les principes susmentionnés de transparence, loyauté et objectivité de l'outil devient essentiel. Une expertise indépendante et régulière devrait pouvoir assurer que les « moteurs » des intelligences artificielles employés pour l'aide à la décision des juges ne sont pas biaisés. Il n'est pas incongru en effet d'anticiper la mise en œuvre, discrète ou non, de dispositifs de référencement payant (sur le modèle du *Search Engine Advertising* de Google) permettant à certains opérateurs d'accorder moins de poids à des décisions leur étant défavorables. Ces règles semblent impératives pour renforcer la confiance des citoyens dans leurs systèmes judiciaires.
160. A ce propos, la qualité reconnue des meilleures réalisations pourrait être matérialisée par la délivrance d'un label ou d'une certification. Doivent notamment être garanties une totale transparence et une parfaite loyauté des modalités de traitement de l'information, tant en direction des professionnels que des citoyens, pour éviter que des exemples comme celui précédemment mentionné de l'algorithme COMPAS puissent se reproduire. En effet, une implication très étroite des professionnels de la justice est essentielle pour pouvoir bien mesurer les enjeux et l'impact de ces applications sur les systèmes judiciaires.
161. L'ensemble des experts participant au développement des IA, notamment les chercheurs, ingénieurs et développeurs informatiques, se trouvent aujourd'hui en charges de responsabilités tout à fait exceptionnelles et inédites. Leur formation pourrait être assortie d'un renforcement encore plus important des « humanités ». L'exemple de certaines écoles novatrices de développeurs informatiques montre que derrière la volonté de « *hacker le système* » se cache, selon certains observateurs, en réalité un pragmatisme dénué de toute contextualisation de la responsabilité qui pèse désormais sur des techniciens aux pouvoirs quasi-démiurgiques ». Le serment d'Hippocrate a certainement ses limites en médecine mais ritualise la prise de responsabilités et offre un cadre éthique.
162. Enfin, la cyberéthique doit s'accompagner d'une formation à large échelle des acteurs, depuis les concepteurs des algorithmes et les legaltechs jusqu'à leurs utilisateurs. De nouvelles humanités, transdisciplinaires, devraient être partagées entre tous pour l'IA devienne le vecteur de développements positifs pour l'humanité et renforce les garanties de la démocratie et de l'État de droit.

Annexe II : Quelles utilisations pour l'IA dans les systèmes judiciaires européens ?

Cette Annexe de la Charte passe en revue différentes utilisations de l'IA dans les systèmes européens et encourage à un degré différent leur application à la lumière des principes et des valeurs énoncés dans la Charte éthique.

L'utilisation de l'apprentissage automatisé (*Machine Learning*) pour constituer des moteurs de recherche valorisant le patrimoine jurisprudentiel est une opportunité à développer pour l'ensemble des professionnels du droit. D'autres applications (constitution de barèmes, appui à des mesures alternatives de règlements de litige) sont à considérer en prenant des précautions (qualité de la source des données notamment et non traitement de masse de l'ensemble du contentieux concerné). Les autres applications (« justice dite prédictive ») seraient à renvoyer au champ de la recherche et d'un développement ultérieur (en consultation avec les professionnels de la justice afin d'assurer une pleine correspondance aux besoins), avant d'en envisager une exploitation publique sérieuse.

En matière pénale, la question apparaît comme extrêmement sensible mais ne doit pas être ignorée. A la lumière des nombreux questionnements existants quant à leur compatibilité avec un certain nombre de droits fondamentaux, l'utilisation d'algorithmes de calcul de potentiels risques de récidive d'un individu traduit en justice devrait être envisagée avec les plus extrêmes réserves. Le traitement de données quantitatives globales pour la prévention de la délinquance est en revanche une piste à approfondir avec ces techniques nouvelles, en prenant en compte les biais connus (effet performatif, qualité des données). De même, l'utilisation d'algorithmes pour mieux faire le lien entre l'offre de travail d'intérêt général et la personnalité d'un individu constituerait peut-être un facteur d'efficacité de ce type de mesure.

➤ Utilisations à encourager

- **Valorisation du patrimoine jurisprudentiel** : les techniques de *Machine Learning* ont été de plus en plus utilisées dans le domaine du traitement du langage naturel au cours des dernières années (ce qui comprend les efforts initiaux de compréhension du langage naturel) et constituent un atout considérable pour bâtir des solutions de recherche complémentaires aux actuels mots-clés ou recherche « plein texte ». Ces outils pourraient lier plusieurs sources (Constitution et conventions, lois, jurisprudence, doctrine). Des techniques de visualisation de données pourraient illustrer le résultat des recherches.
- **Accès au droit** : sans remplacer une intervention humaine, des chatbots pourraient être mis en place pour faciliter l'accès, en langage naturel, aux différentes sources d'information existantes. Des modèles d'actes (saisine d'un tribunal, contrat de bail par exemple) pourraient aussi être générés en ligne.
- **Construction de nouveaux outils de pilotage** : l'utilisation des techniques de datascience et d'intelligence artificielle sur les données d'activité des juridictions peut contribuer à améliorer l'efficacité de la justice en permettant notamment de procéder à des évaluations quantitatives et qualitatives et à construire des projections (anticipation des moyens humains et budgétaires). Des indicateurs clés de performance pourraient être constitués sur cette base. L'implication des professionnels de justice, et notamment des juges, dans la mise en place de ces outils est recommandée, tant sur le plan de l'appropriation de ceux-ci que sur celui de l'analyse des résultats et leur pondération avec des éléments relatifs aux spécificités de la juridiction ou de qualité de la justice (par exemple, la nécessité de préserver l'accès à la justice).

➤ **Utilisations à envisager avec de fortes précautions méthodologiques**

- **Aide à la construction de barèmes/échelles dans certains contentieux civils** : l'analyse de la masse totale des décisions judiciaires n'a statistiquement pas grand sens si l'on n'identifie pas tous les facteurs causatifs (explicites et implicites dans les décisions). Savoir que la moyenne de la prestation compensatoire dans une certaine zone géographique est plus haute que dans une autre est explicable non du fait du comportement des juges, mais au vu de la caractéristique du territoire. Le *Machine Learning* peut donc être utile pour identifier des décisions (cf valorisation du patrimoine jurisprudentiel supra) mais le traitement automatisé ne saurait seul produire du sens. La construction d'un échantillon pertinent de décisions à traiter (cf technique de sondages par exemple) est un préalable indispensable.
- **Appui à des mesures alternatives de règlement de litiges en matière civile** : dans certains pays européens, des outils de justice dite « prédictive » sont utilisés par des compagnies d'assurance pour évaluer les chances de succès d'un contentieux et détourner le justiciable vers un autre mode de règlement de son litige lorsque les chances de succès sont considérées faibles. Par ailleurs, certains systèmes à l'étranger proposent des montants d'indemnisation sans réelle transparence des règles de calcul. Toutefois, l'évaluation de ces systèmes ne peut être considérée comme impartiale et fiable (cf point sur les techniques de *Machine Learning*) : on fait prendre une décision au citoyen sur des bases tronquées. Dans d'autres cas encore, une mesure alternative de règlement de litige peut être conseillée au justiciable par un agent relationnel (chatbot) après un examen préliminaire des critères renseignés par le justiciable lui-même, lorsque celui-ci se rend sur le site d'une juridiction ou recherche sur le web une information juridique. Cet agent relationnel peut, le cas échéant, également recommander au justiciable de demander le conseil d'un avocat ou d'un service de médiation. Dans toutes ces hypothèses, la présence d'un tiers formé (médiateur s'appuyant non seulement sur des techniques mais peut-être des barèmes comme calculé plus haut, ou un avocat) semble être la solution la plus appropriée à ce stade.
- **Règlement des litiges en ligne** : lorsqu'il se rend sur une plateforme de résolution de litige en ligne, le justiciable devrait être informé d'une manière claire et compréhensible si le traitement de son litige se fait de manière entièrement automatisée ou avec l'implication d'un médiateur ou d'un arbitre. Par ailleurs, l'information au justiciable devrait aussi être loyale, et éviter de lui donner l'impression de saisir un tribunal (à cet égard, on note que la dénomination « tribunal en ligne » est souvent utilisée pour ce type de plateformes, alors que techniquement elles visent à fournir des services de règlement alternatif des litiges). Il s'agit de deux éléments essentiels pour permettre au justiciable de faire un choix éclairé et décider éventuellement de s'y opposer et des recourir à un vrai tribunal, au sens de l'article 6 de la CEDH. Par ailleurs, eu égard aux exigences des articles 6 et 13 de la CEDH, il faudrait toujours envisager des formes de contrôle de la procédure de règlement des litiges en ligne et de son issue par les juridictions de l'Etat, surtout lorsque le justiciable a consenti au règlement des litiges en ligne en forme entièrement automatisée.
- **Utilisation des algorithmes en matière d'enquête pénale afin d'identifier des lieux de commission d'infractions** : ce genre d'application pourrait non-seulement concerner la police mais également les parquets dans le cadre des instances de prévention de la délinquance auxquelles elles participent. Là encore des systèmes ont été utilisés aux Etats-Unis afin de diriger en temps réel des patrouilles de police vers des lieux possibles de commission d'infraction. Ce type d'approche quantitative est toutefois susceptible de générer un fort effet « performatif » (en un lieu, on a plus de chance de découvrir une infraction et l'on autoalimente ainsi le système). L'analyse criminelle par des approches couplées entre des systèmes d'information géographique (SIG) et des masses de données des procédures pourraient être mieux partagées avec les parquets et bénéficier certainement d'un fort apport du *Machine Learning*. Des unités de lutte contre le blanchiment utilisent déjà des systèmes « prédictifs » pour identifier des flux financiers suspects, mais s'agissant d'information quantitative (monétaire), les machines sont plus à même de produire des analyses fiables. La recherche devrait aussi avoir un meilleur accès à ces données pour produire des études pertinentes à destination des décideurs publics.

➤ **Utilisations à envisager au terme de travaux scientifiques complémentaires**

- **Profilage des juges** : la quantification de l'activité d'un juge rendra moins compte de ses possibles biais de raisonnement que de facteurs externes ayant pesé sur sa décision. L'activité juridictionnelle dans un territoire paupérisé ne produit pas les mêmes résultats que dans un autre territoire, et cela indépendamment de la personnalité du juge. Lorsque la décision est rendue de manière collégiale et sans possibilité pour un juge d'exprimer une opinion divergente, il est vain de profiler chacun des juges de la chambre. Offrir en revanche aux juges une évaluation quantitative et qualitative plus fine de leurs activités, grâce à de nouveaux outils, mais dans un but purement informatif d'aide à la décision et à leur usage exclusif pourrait en revanche être encouragé.
- **Anticipation des décisions de justice** : le seul traitement statistique de groupes lexicaux révèle la fréquence de l'utilisation de certains groupes de mots mais n'identifie par les causes réelles d'une décision car il ne procède pas d'un raisonnement juridique (cf. étude conduite sur les décisions de la CEDH par l'*University College of London* qui obtient des résultats meilleurs sur la partie fait que sur la partie analyse du droit). Des systèmes hybrides, fondés sur la construction de modèles mathématiques censés représenter la diversité de raisonnement des juges, ne sont pas plus efficaces car ils restent limités aux biais de l'échantillon de données qu'ils ont traités et nécessitent d'être totalement réécrits en cas de modification de la loi ou de revirement de jurisprudence.

➤ **Utilisations à envisager avec les plus extrêmes réserves**

- **Utilisation des algorithmes en matière pénale afin de profiler les individus** : les expériences étrangères (COMPAS aux Etats-Unis et HART en Grande-Bretagne) ont été dénoncées par des ONG (cf travaux de *ProPublica* aux Etats-Unis ou de *Big Brother Watch* en Grande Bretagne). En raison des limites des méthodes employées, cette approche purement statistique a conduit à des résultats erronés : au constat que certains individus afro-américains seraient plus souvent impliqués dans des faits criminels, il en a été déduit un facteur de risque plus élevé pour toute la population afro-américaine. Donc, même pour des faits mineurs, ces systèmes ont pondéré négativement des accusés afro-américains et ont conduit à un injuste alourdissement du quantum de leurs peines. Une orientation différente, plus respectueuse des standards européens concernant la sanction pénale, et qui doit offrir des chances de réhabilitation et de réintégration à tout individu, doit remplacer cette approche aux effets discriminatoires et déterministes. Si des systèmes algorithmiques arrivent à contribuer à une meilleure collecte d'information au profit des services de probation par exemple, et à leur offrir de manière plus rapide les informations pertinentes pour un traitement respectueux des droits de l'Homme, de telles initiatives peuvent être utiles (notamment pour les procédures de jugement rapide). Toute autre utilisation est en revanche sujette à des biais qui heurteront de nombreux principes fondamentaux nationaux et supranationaux.
- **« Norme issue du nombre »** il ne s'agit pas seulement de produire des barèmes, ce qui pourrait être légitime, mais de fournir à chaque juge le contenu des décisions produite par l'ensemble des autres juges et de prétendre enfermer son futur choix dans la masse de ces « précédents ». Or le nombre ne saurait ajouter à la loi ou s'y substituer. Pour les raisons évoquées ci-dessus, la masse ne fait pas nécessairement sens. L'étude de la CEPEJ a aussi souligné les risques existants en matière de cristallisation de la jurisprudence et de ses effets potentiellement négatifs sur l'impartialité et l'indépendance des juges.

Annexe III : Glossaire

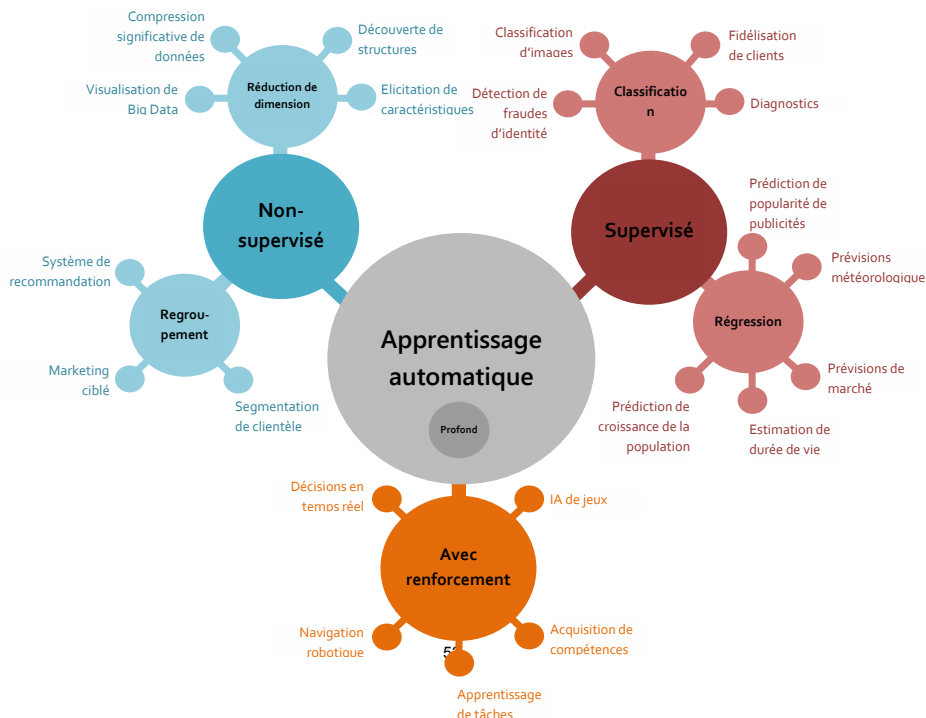
Le présent glossaire vise à définir les termes employés par la charte éthique et le document d'études. Il a été préféré une acception étroite à l'ensemble du vocabulaire employé. L'ensemble des documents doivent être lus et compris à la lumière de ces définitions

A

ALGORITHME Suite finie de règles formelles (opérations logiques, instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Cette suite peut être l'objet d'un processus automatisé d'exécution et s'appuyer sur des modèles conçus par le biais d'apprentissage machine.

APPRENTISSAGE MACHINE (*machine learning*) L'apprentissage machine permet de construire un modèle mathématique à partir de données, en incluant un grand nombre de variables qui ne sont pas connues à l'avance. Les paramètres sont configurés au fur et à mesure lors d'une phase d'apprentissage, qui utilise des jeux de données d'entraînement pour trouver des liens et les classifie. Les différentes méthodes d'apprentissage machine sont choisies par les concepteurs en fonction de la nature des tâches à accomplir. Ces méthodes sont habituellement classées en 3 catégories : apprentissage supervisé par un humain, apprentissage non supervisé et apprentissage par renforcement. Ces 3 catégories regroupent différentes méthodes dont les réseaux de neurones, l'apprentissage profond etc.

Le graphique ci-dessous illustre les différentes catégories d'apprentissage machine :



ANONYMISATION Processus consistant à traiter des données à caractère personnel afin d'empêcher totalement et de manière irréversible l'identification d'une personne physique ou morale. L'anonymisation suppose donc qu'il n'y ait plus aucun lien possible entre l'information concernée et la personne à laquelle elle se rattache. L'identification devient alors totalement impossible¹. Les principes relatifs à la protection des données relatives à une personne physique identifiée ou identifiable n'ont donc lieu à s'appliquer si ces données sont totalement anonymisées.

B

BASE DE DONNÉES Une base de données est un « conteneur » stockant des données telles que des chiffres, des dates ou des mots, pouvant être retraités par des moyens informatiques pour produire une information ; par exemple, des chiffres et des noms assemblés et triés pour former un annuaire.

BIG DATA (mégadonnées, grand jeux de données) Le terme « big data » désigne un grand ensemble de données, de source hétérogène (open data, données propriétaires, données achetées commercialement). Pour les données issues de l'activité judiciaire, le *big data* pourrait être la réunion des données statistiques, des traces de connexion dans les logiciels des métiers (logs des applications), des bases de données des décisions des tribunaux etc.

C

CHATBOT (agent conversationnel) Agent conversationnel qui dialogue avec son utilisateur (par exemple : les robots empathiques à disposition de malades, ou les services de conversation automatisés dans la relation au client)².

D

DATAMINING (Analyse et fouille de données) Le *datamining* permet d'analyser un grand volume de données et d'en faire ressortir des modèles, des corrélations, des tendances.

DEEP LEARNING (Apprentissage profond) Voir *Apprentissage machine* et *Neurones*

DONNÉES Représentation d'une information en vue d'un traitement automatique. Lorsqu'on dit que les algorithmes peuvent « s'appliquer » à des réalités les plus diverses dans le monde du droit ou ailleurs, on présuppose le caractère numérisable de toute réalité sous forme de « données ». Mais on voit bien dès le cas de la physique, que rien ne nous dit que les processus physiques peuvent être traduits adéquatement en termes de « données » (et s'intégrer au cycle entrée / sortie des algorithmes). Si c'est déjà le cas en physique, il n'y a pas de raison que ce ne soit pas non plus le cas dans les rapports sociaux. Il faut donc être prudent avec l'idée de « données » qui suppose toujours que la réalité que l'on cherche à décrire possède un format tel qu'il soit naturellement traitable algorithmiquement.

¹ Avis 05/2014 du G29 sur les techniques d'anonymisation. Voir aussi considérant n° 26 du règlement européen (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016.

² Rapport Commission Nationale de l'Informatique et des Libertés, Décembre 2017 : Comment permettre à l'homme de garder la main ? les enjeux éthiques des algorithmes et de l'intelligence artificielle.

DONNÉES À CARACTÈRE PERSONNEL Toute information concernant une personne physique identifiée ou identifiable (dite « personne concernée »), directement ou indirectement. Parmi celles-ci, les données sensibles concernent les données génétiques, les données biométriques identifiant un individu de façon unique, les données concernant des infractions, des procédures et des condamnations pénales et des mesures de sûreté connexes ainsi que toutes données pour les informations qu'elles révèlent sur l'origine raciale ou ethnique, les opinions politiques, l'appartenance syndicale, les convictions religieuses ou autres convictions, la santé ou la vie sexuelle.



INTELLIGENCE ARTIFICIELLE (IA) Ensemble de sciences, théories et techniques dont le but est de reproduire par une machine des capacités cognitives d'un être humain. Les développements actuels visent à pouvoir confier à une machine des tâches complexes auparavant déléguées à un humain.

Le terme d'intelligence artificielle est toutefois critiqué par les experts qui réalisent une distinction entre des IA qualifiées de « fortes » (en capacité de contextualiser des problèmes spécialisés très différents de manière totalement autonome) et des IA « faibles » ou « modérées » (pourtant extrêmement performantes dans leur domaine d'entraînement). Les IA « fortes » nécessiteraient, selon certains experts, des progrès significatifs en recherche fondamentale pour être en capacité de modéliser le monde en son entier et non de simples améliorations de performance des systèmes existants.

Les outils recensés dans le cadre de ce document sont développés en utilisant des méthodes d'apprentissage machine, donc des IA « faibles ».



JUSTICE dite « PRÉDICTIVE » On entend par « justice prédictive » l'analyse de grandes masses de décisions de justice par des technologies d'intelligence artificielle afin de construire, pour certains types de contentieux spécialisés, des prévisions sur l'issue des litiges (montant des indemnités de licenciement ou de la pension alimentaire par exemple).

Le terme « prédictif » utilisé par les *legaltechs* provient des sciences (essentiellement statistiques) qui permettent de « prédire » des résultats futurs par une analyse inductive. Un traitement est fait à partir de l'analyse des décisions judiciaires qui vise à identifier des corrélations entre des données en entrée (les critères de la loi, les faits de l'affaire, la motivation) et des données en sortie (dispositif du jugement tel que le montant d'une indemnité). Les corrélations jugées comme pertinentes permettent de bâtir des modèles qui, utilisés avec de nouvelles données d'entrée (nouveaux faits ou précisions nouvelles en paramètre, telles que par exemple la durée d'une relation contractuelle), produisent selon leurs concepteurs une prévision de décision (fourchette d'indemnisation par exemple).

Cette approche est contestée sur la forme et le fond par certains auteurs. Ceux-ci avancent en effet que, de manière générale, la modélisation mathématique de certains phénomènes sociaux n'est pas une tâche comparable à d'autres activités plus aisément quantifiables (isoler les facteurs réellement causatifs d'une décision judiciaire est infiniment plus complexe que de jouer au jeu de Go ou reconnaître une image par exemple) : les risques de corrélations fallacieuses seraient bien plus importants. De plus, en théorie du droit, deux décisions contradictoires peuvent s'avérer valide si le raisonnement juridique n'est pas contestable. Dresser des prévisions, dès lors, serait un exercice à la vocation purement informative sans aucune prétention prescriptive.

L

LEGALTECH Entreprises exploitant les technologies de l'information dans le domaine du droit afin de proposer des services juridiques innovants. Ces entreprises sont des startups spécialisées dans le domaine juridique. D'autres termes dérivés des secteurs d'activités sont aussi apparus comme les « *fintech* » pour les startups déployant des services financiers, les « *medtechs* » dans le domaine médical, etc.

LOGICIEL EN OPEN SOURCE Logiciel dont le code source est ouvert. Le logiciel peut ainsi être utilisé, modifié et redistribué librement.

M

MACHINE LEARNING (*apprentissage machine ou apprentissage automatisé*)
Voir *apprentissage machine*.

MÉTADONNÉES Données qui permettent de définir, de contextualiser ou de caractériser d'autres données. Dans la plupart de ses usages informatiques, le préfixe méta signifie « définition ou description de référence ». Les métadonnées synthétisent des informations élémentaires sur les données, elles facilitent la recherche et la manipulation d'instances de données particulières. *L'auteur, la date de création, la date de modification et la taille du fichier* en sont des exemples. Les métadonnées et leur corollaire, le filtrage des données, aident à localiser un document spécifique.

N

NEURONES / RESEAU DE NEURONES Les réseaux neuronaux sont des systèmes informatiques vaguement inspirés des réseaux neuronaux biologiques qui constituent le cerveau des animaux. Ces systèmes « apprennent » à effectuer des tâches en considérant des exemples, généralement sans être programmés avec des règles spécifiques. Par exemple, dans la reconnaissance d'images, ils peuvent apprendre à identifier les images qui contiennent des chats en analysant des images d'exemple qui ont été étiquetées manuellement comme « chat » ou « sans chat » et en utilisant les résultats pour identifier les chats dans d'autres images. Ils le font sans aucune connaissance préalable des chats, par exemple, qu'ils ont de la fourrure, des queues, des moustaches et des visages de chats. Au lieu de cela, ils génèrent automatiquement des caractéristiques d'identification à partir du matériel d'apprentissage qu'ils traitent.

Un NN est basé sur un ensemble d'unités ou de nœuds connectés appelés neurones artificiels qui modélisent de façon lâche les neurones dans un cerveau biologique. Chaque connexion, comme les synapses d'un cerveau biologique, peut transmettre un signal d'un neurone artificiel à un autre. Un neurone artificiel qui reçoit un signal peut le traiter puis signaler d'autres neurones artificiels qui lui sont connectés.

Le but initial de l'approche NN était de résoudre les problèmes de la même manière qu'un cerveau humain le ferait. Cependant, avec le temps, l'attention s'est portée sur l'exécution de tâches spécifiques, ce qui a entraîné des écarts par rapport à la biologie. Les réseaux neuronaux artificiels ont été utilisés pour une variété de tâches, y compris la vision par ordinateur, la reconnaissance vocale, la traduction automatique, le filtrage des réseaux sociaux, les jeux de société et les jeux vidéo et le diagnostic médical.

O

OPEN DATA Le terme désigne la mise à disposition publique, par téléchargement, de bases de données structurées. Ces données sont ré-employables de manière non-onéreuse dans les conditions d'une licence spécifique, pouvant notamment préciser ou prohiber certaines finalités de réemploi.

L'*open data* n'est pas à confondre avec de l'information publique unitaire disponible sur des sites internet, dont l'intégralité de la base n'est pas téléchargeable (par exemple les bases de données des décisions des tribunaux). Il ne se substitue pas aux modes de publicité obligatoire de certaines mesures ou décisions administratives ou judiciaires déjà prévues par certaines lois ou règlements.

Enfin, une confusion est parfois réalisée entre les données (*open data* à proprement parler) et leurs moyens de traitement (apprentissage machine, *data science*) pour différentes finalités (moteurs de recherche, aide à la rédaction d'actes, analyse des tendances décisionnelles, anticipation des décisions de justice).

P

PROFILAGE Une technique de traitement automatisé des données qui consiste à appliquer un « profil » à une personne physique, notamment afin de prendre des décisions à son sujet ou d'analyser ou de prévoir ses préférences, comportements et attitudes personnels.

PSEUDONYMISATION Traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable³.

S

SCIENCES DES DONNEES Regroupement large des mathématiques, des statistiques, des probabilités, de l'informatique, de la visualisation des données afin d'extraire des connaissances d'un ensemble hétérogène de données (images, son, texte, données génomiques, liens dans les réseaux sociaux, mesures physiques, etc.).

Les méthodes et outils issus de l'intelligence artificielle entrent dans cette famille.

SYSTEME EXPERT Il s'agit de l'une des voies tentant d'aboutir à l'[intelligence artificielle](#). Un système expert est un outil capable de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier. Plus précisément, il s'agit d'un [logiciel](#) capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connues. Il se compose de 3 parties :

- une base de faits ;
- une base de règles ;
- un moteur d'inférence.

Le moteur d'inférence est capable d'utiliser faits et règles pour produire de nouveaux faits, jusqu'à parvenir à la réponse à la question experte posée.

La plupart des systèmes experts existants reposent sur des mécanismes de logique formelle (logique aristotélicienne) et utilisent le raisonnement déductif.

³ Article 4 du Règlement européen (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016.

T

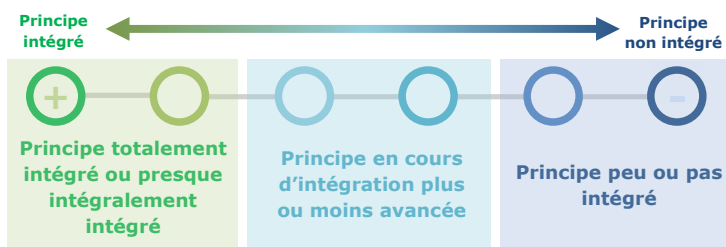
TRAITEMENT DE DONNÉES À CARACTÈRE PERSONNEL Selon l'Article 2 de la Convention 108 modernisée, toute opération ou ensemble d'opérations effectuées sur des données à caractère personnel, telles que la collecte, l'enregistrement, la conservation, la modification, l'extraction, la communication, la mise à disposition, l'effacement ou la destruction des données, ou l'application d'opérations logiques et/ou arithmétiques à ces données.

Annexe IV : Checklist d'intégration des principes de la Charte dans vos traitements

Afin d'évaluer la compatibilité de votre traitement avec la Charte, une échelle d'autoévaluation vous est proposée pour chacun des principes d'action énoncés.

Pour chaque principe, cochez l'une des 6 cases vous paraissant adaptée au contexte de votre traitement.

La case la plus à gauche indique une intégration totale, la case la plus à droite l'absence d'intégration.



Additionnez en bas de la feuille d'évaluation (ligne « total ») le nombre de cases cochées. Le nombre majoritaire vous indique le degré de compatibilité de votre traitement avec la charte.

Cette évaluation est bien entendu purement indicative et n'équivaut pas à délivrer un label ou une certification.

Checklist d'évaluation de votre traitement

1. **Principe de respect des droits fondamentaux :**
Assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux, notamment le droit la protection des données personnelles



2. **Principe de prévention :**
Prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus



3. **Principe de qualité et sécurité :**
En ce qui concerne le traitement de décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles, avec des modèles conçus d'une manière disciplinaire, dans un environnement technologique sécurisé



4. **Principe de transparence, de neutralité et d'intégrité intellectuelle :**
Rendre accessible et compréhensible les méthodologies de traitement des données, autoriser les audits externes



5. **Principe de maîtrise par l'utilisateur :**
Bannir une approche prescriptive et rendre l'utilisateur un acteur éclairé et maître de ses choix



Compatible avec la charte

Mesures à prendre pour être compatible

Non compatible avec la charte



PODER JUDICIÁRIO DO
ESTADO DE RONDÔNIA



**APOIO
INSTITUCIONAL**