UNIVERSIDADE CATÓLICA PORTUGUESA

Viabilidade da aplicação de sistemas de inteligência artificial aos atos administrativos discricionários

Jaime Paulino Maia e Silva

Mestrado em Direito

Faculdade de Direito | Escola do Porto

2023

UNIVERSIDADE CATÓLICA PORTUGUESA

Viabilidade da aplicaçãode sistemas de inteligência artificial aos atos administrativos discricionários

Jaime Paulino Maia e Silva

Orientador: Professor Pedro Miguel dos Santos Silva Cerqueira Gomes

Mestrado em Direito

Faculdade de Direito | Escola do Porto

Maio, 2023

À Maria José, minha esposa, pelo amor incondicional,

e pela resiliência, apoio e ajuda inexcedíveis, e ainda, por tudo aquilo que não sou capaz de exprimir desta forma. Muito obrigado.

Ao Professor Pedro Cerqueira Gomes, pela visão e conhecimento nestas "novas avenidas" do Direito Administrativo,

pela energia e entusiasmo contagiantes, pela excelente relação e, sobretudo, por ter sido uma fonte de inspiração, o meu sincero obrigado!

Obrigado à Mafalda Lopes, por esta ajuda final, tão preciosa!

Não tenhamos pressa. Mas não percamos tempo.

José Saramago

Resumo

Os algoritmos têm evoluído exponencialmente nas últimas décadas e a sua implementação em todos os setores de atividade resulta simultaneamente de uma vontade – um ato voluntário - e de uma inevitabilidade. Porém quanto mais evoluídos e potentes se tornam, quantos mais dados conseguem analisar e quanto mais conseguem trabalhar com bases de dados não estruturadas, maiores são os seus efeitos resultantes das suas previsões e decisões. Existem muitos riscos associados a bases de dados enviesadas ou não representativas da sociedade, uma vez que desfavorecem grupos de risco ou minorias, provocando a discriminação, o enviesamento decisório, exponenciados pela escalabilidade e massificação dos outputs produzidos por sistemas automatizados e não dando garantias de equidade.

Por outro lado, existem questões associadas à opacidade algorítmica ou efeito "black-box", em que existe falta de transparência, de explicabilidade e interpretabilidade inteligível por todos, mas fundamentalmente pelos utilizadores intermédios ou pelos finais, das decisões e do porquê das mesmas, num direito à fundamentação.

Se este contexto é já de si complexo, a sua aplicação ao campo da Administração Pública, imbuída dos seus poderes de "autonomia valorativa", a procura de respostas à questão da viabilidade da aplicação de sistemas de IA aos atos administrativos discricionários, parece ser um desafio excessivo.

Para abordar tal desafio, propusemos uma metodologia jurídico-administrativa própria, estruturada em 4 dimensões, de abordagem ao tema da automatização dos atos administrativos por níveis de decisão, de acordo com o grau de complexidade e de discricionariedade.

Essa metodologia assenta num processo de automatização colaborativa e corresponsável, entre o agente público e os sistemas de IA, através de um modelo de inteligência colaborativa homem-máquina, suportado por aquilo que de mais recente existe no campo da IA - que não apenas os algoritmos de autoaprendizagem Machine e Deep Learning, mas outros bem mais evoluídos e responsivos aos requisitos que se impõem na implementação destas ferramentas no âmbito do Direito e, em particular do Direito Administrativo, em termos transparência, da IA explicável, da IA causal, de algoritmos interativos, ou dos algoritmos de computação cognitiva.

Palavras-chave: inteligência artificial, inteligência artificial causal, aprendizagem profunda, aprendizagem automática, homem no circuito, viés algorítmico, explicabilidade, discricionariedade, transparência, ato administrativo discricionário

Abstract

In recent decades algorithms have evolved exponentially and their implementation in all sectors of activity results simultaneously from a will – a voluntary act – and from an inevitability.

However, the more evolved and powerful they become, the more data they can analyze and the more they can work with unstructured databases, the greater the effects resulting from their forecasts and decisions. There are many risks associated with databases that are biased or not representative of society, since they disadvantage risk groups or minorities, causing discrimination, decision-making bias, exponentiated by the scalability and massification of outputs produced by automated systems and not giving guarantees of equity.

On the other hand, there are issues associated with algorithmic opacity or the "black-box" effect, in which there is a lack of transparency, explainability and intelligible interpretability by all, but fundamentally by intermediate or end users, of decisions and the reason for them, a right to state reasons.

To address this challenge, we propose our own legal-administrative methodology, structured in 4 dimensions, to approach the topic of automating administrative acts by decision levels, according to the degree of complexity and discretion.

This methodology is based on a process of collaborative and co-responsible automation, between the public agent and the AI systems, through a model of collaborative human-machine intelligence, supported by the latest in the field of AI - not just algorithms Machine and Deep Learning self-learning, but others that are much more evolved and responsive to the requirements that are imposed in the implementation of these tools in the field of Law and, in particular, Administrative Law, in terms of transparency, explainable AI, causal AI, interactive algorithms , or cognitive computing algorithms.

Keywords: discretionary administrative act, artificial intelligence, causal artificial intelligence, deep learning, machine learning, human-in-the-loop, algorithmic bias, explainability, accountability, discretionary, transparency, discretionary administrative act.

Sumário

Lista de siglas e abreviaturas

AA – Ato Administrativo

AAV – Ato Administrativo Vinculado

AAD - Ato Administrativo Discricionário

ApAt - Autoaprendizagem ou aprendizagem automática

AP – Administração Pública

CC – Código Civil

CpCg - Computação Cognitiva

CRP – Constituição da República Portuguesa

DL - Deep Learning

IA - Inteligência Artificial

IAc - Inteligência Artificial Causal

IAexp - Inteligência Artificial Explicável

SIA – Sistema(s) de Inteligência Artificial

ML - Machine Learning

MLa - Machine Learning Autónoma

MLi - Machine Learning Interativa

RGPD - Regulamento Geral de Proteção de Dados

UE – União Europeia

Introdução

Os sistemas de inteligência artificial (SIA) são um tema omnipresente e incontornável, e por maioria de razão, também no Direito Administrativo e na Administração Pública (AP). A questão de sabermos da viabilidade da aplicação dos SIA aos atos administrativos discricionários (AAD), (uma vez que a aplicação aos atos administrativos vinculados (AAV) de natureza binária, as possibilidades são hoje já uma evidência) é oportuna, faz sentido e é urgente.

Seguimos a via do "como", para chegar ao "sim" ou ao "não" da resposta. Estruturamos a reflexão em duas partes.

Cap. I, com notas relevantes para reforço da delimitação de um quadro jurídico de enquadramento e suporte ao tema;

Cap. III, ensaiamos a conceção de um modelo, próprio, de abordagem ao "como" da implementação de SIA aos atos administrativos, desde os AAV simples aos discricionários complexos.

A meio da viagem fizemos uma breve paragem – no cap. II, trazendo à colação alguns afloramentos aos métodos de raciocínio das duas ciências – Direito e Computação, à procura de elementos convergentes e sinérgicos, de reforço da confiança para seguirmos em frente!

Capítulo 1. Construção do quadro jurídico

Considerações introdutórias

Hoje estamos num patamar já muito avançado, distante dos algoritmos "if this … then that", no mundo da aprendizagem automática, dos Machine Learning (ML), dos Deep Learning (DL) em estádios da "inteligência da máquina" (rede avançada, treinada para construir modelos ad hoc para aprenderem sobre dados customizados) e a caminho de uma superinteligência da "consciência da máquina" (autoaprendizagem cognitiva). Por outro lado, existe uma crescente prevalência da tomada de decisão algorítmica pelas autoridades públicas, com impacto nos cidadãos.

Posto o que, se os benefícios são enormes, os perigos, os requisitos e todos os cuidados a ter são igualmente significativos.

É por aqui que começaremos analisando os riscos ao nível da opacidade, da discriminação; e abordando a dificuldade em assegurar as condições bastantes ao nível da transparência, explicabilidade, compreensão das decisões e prestação de contas, … para uma Inteligência Artificial (IA) de confiança.

1. Riscos associados à Inteligência Artificial

1.1. Opacidade, o efeito “caixa-preta”

A IA atual não é mais uma abordagem puramente determinista, em que decisão, regras e equações matemáticas foram definidas por seres humanos e implementadas para serem processadas automaticamente. Hoje, estas regras são extraídas diretamente das bases de dados. Consequentemente, a decisão automatizada perdeu a sua interpretabilidade inerente, uma vez que as regras podem ser facilmente ocultadas pela complexidade dos algoritmos - quanto mais eficiente for o algoritmo, mais opaco ele será (Vérine, 2019; Pilving, 2020). Acresce ainda que o tamanho gigantesco dos megadados trabalhados atualmente, dificulta, se não mesmo impede, a sua análise e compreensão por um cérebro humano.

Então temos a "opacidade algorítmica" ou o efeito “caixa preta”, que se traduz no facto de não podermos explicar a lógica e os motivos de uma decisão algorítmica, porque apesar dos dados de entrada e saída serem conhecidos, o seu funcionamento interno específico, não é determinavel ou entendível. A representação interna dos dados é abstrata e complexa de decifrar, assim como as regras codificadas durante a fase de treino ou de construção do algoritmo.

Como resultado o processo de tomada de decisão de um SIA pode ser difícil de entender ou impossível de avaliar, pelos programadores e engenheiros de sistemas e muito menos compreensível para utilizadores finais. Em muitos casos, é praticamente impossível determinar como ou porquê um determinado resultado foi alcançado, com o risco de criarmos uma “black box society”, [1] [2] que importa conhecer e banir dos SIA. Se assim não for estaremos perante a violação de um conjunto de princípios, entre eles o da igualdade (art.º 13 da CRP), com consequências a nível da transparência, responsabilização e justiça (Huggins, 2020).

[1] Pasquale, 2015; Brown, 2016

[2] Biased algorithms – algoritmos discriminatórios que contribuem para agravar e perpetuar problemas sociais, raciais ou outros dos grupos minoritários, ver Brown, 2016; Pasquale, 2015); ou, como efeitos colaterais teremos os inductive bias (Learning bias), em relação às classes maioritárias, se os dados de treino não forem balanceados, resultando num desempenho empobrecido no reconhecimento de classes minoritárias, Dong, 2019.

A transparência e a responsabilidade das decisões, em termos dos métodos utilizados e, consequentemente, dos resultados alcançados ficam postas em causa. A incapacidade de analisar as razões por trás das recomendações do algoritmo pode prejudicar aqueles afetados por tais decisões. A opacidade pode destruir o sentido de justiça e confiança dos visados.[3]

Em síntese, a implementação de SIA pode ser um risco para o procedimento administrativo: a IA determinará qual é a parte processual que tem mais força (AP ou cidadão) de forma opaca e unilateral, correndo o risco de que vença a "lei do mais forte" (Santos, 2022).

Um SIA que toma decisões em nome do Estado não pode ser uma "caixa preta", mas pautar-se por praticas responsáveis, devendo aderir aos seguintes requisitos: explicabilidade – compreender o racional por detrás de cada decisão; transparência – perceber o modelo de decisão algorítmica; e probabilidade – certeza matemática por trás da decisão; e seguindo cinco princípios essenciais: equidade, transparência, prestação de contas; processos de verificação e auditoria contínuas e supervisão humana. Abordaremos algumas destas preocupações ao longo do trabalho.

1.2. Enviesamento & Discriminação

Enviesamento diz respeito à utilização ou escolha de dados ou de resultados algorítmicos que resultem, direta ou indiretamente, num efeito discriminatório ilícito ou antiético sobre um indivíduo, ou quando os dados selecionados não sejam representativos[4].

1.2.1. Grau de neutralidade algorítmica:

Existe uma corrente, que descarta o enviesamento, usando o argumento da neutralidade como forma de rejeitar a responsabilidade (Boyd, 2016). Contudo, muitos outros trabalhos demonstraram como SIA em áreas como o policiamento (Ensign, 2018), a justiça (Chouldechova, 2017), a moderação de conteúdo online (Binns, 2017), entre

[3] Entre outros: Super, 2005; Grimmelmann, 2005; Citron, 2008; Pasquale, 2015; Hogan-Doran, 2017; Desai, 2017; Kroll, 2017; Mulholand, 2019; Deeks, 2019.
[4] Agência dos Direitos Fundamentais da União Europeia (FRA), #BigData: Discrimination in Data-Supported Decision Making, Viena: FRA, 2018.

outros, estão longe de serem neutros, apresentando casos detetáveis de injustiça entre a forma como os diferentes grupos são representados nesses sistemas.

Sabemos hoje que o algoritmo pode ser programado para introduzir inadvertidamente preconceitos, reforçar a discriminação, favorecer uma orientação política ou práticas indesejadas, de forma automática, durante o processo de treino, a partir dos dados que está a operar (Harlow, 2019). E, a um nível mais amplo e grave, coloca-se a questão de saber: a quem a tecnologia capacita ou descapacita ao longo do tempo (Veale, 2019).

1.2.2. Preocupações e dificuldades:

O uso do ML tem suscitado uma série de preocupações, especialmente: quando as suas decisões têm um impacto, pela sua magnitude e escala, muito superior às tomadas por humanos; quando muitos dos dados utilizados para treinar os SIA serem pouco representativos da população em geral, suscitando decisões injustas, que refletem os preconceitos mais amplos da sociedade; e, sobretudo, quando os sistemas fazem previsões que afetam a liberdade, a segurança ou a privacidade dos cidadãos (House of Lords, 2018; Williams, 2020).

Os dados são produzidos por seres humanos, com toda a subjetividade que isso implica. Se as decisões administrativas incluídas nas bases de dados de treino, tiverem um elemento subjetivo e enviesado, o algoritmo não tem capacidade para alterar as suas decisões para que sejam automaticamente justas, imparciais, objetivas e legais; também é importante estar consciente de que o viés pode surgir quando os conjuntos de dados refletem incorretamente a sociedade, mas também pode surgir quando os conjuntos de dados refletem com precisão aspetos injustos da sociedade; ou, ainda quando determinados grupos são ou foram historicamente tratados menos favoravelmente do que outros, isso pode "produzir um modelo" que repete esta diferença de tratamento, esta desvantagem. Como resultado, os SIA podem ser propensos a tomarem decisões que são sistematicamente distorcidas, em vez de agirem de forma imparcial. Tal pode levar a que aqueles que preenchem determinados critérios, sejam tratados de forma menos favorável do que aqueles que não preenchem esses critérios.[5]

[5] House of Lords, 2018; Anastaspoulos, 2018; Cobbe, 2019; Finck, 2019; Brand, 2020.

Em síntese:

Embora a redução do viés seja uma área de pesquisa em ML, ainda não há consenso sobre o que constitui exatamente viés nos SIA, e sobre os meios confiáveis para o identificar ou eliminar. Algumas pesquisas sugerem mesmo que a eliminação pode ser impossível. Havendo quem refira, que não se pode ser neutro; porque somos nós – humanos -que fazemos as coisas. E isso é projetado no que fazemos. A questão é saber como essas preferências se podem tornar explícitas, porque se elas se podem tornar explícitas, então é possível lidar com isso. Só que os preconceitos por norma estão escondidos e não se veem. É a dificuldade em explicitar coisas implícitas (Kleinberg, 2016; Courtland, 2018; House of Lords, 2018).

Contudo, a IA também pode ajudar a corrigir alguns preconceitos: os humanos são tendenciosos; as máquinas não são, a menos que as treinemos para o ser. A IA também pode fazer um bom trabalho na deteção do viés inconsciente.

2. Requisitos Funcionais

2.1. Transparência

A transparência é uma condição prévia e necessária para assegurar a conformidade legal, permitir a responsabilização e um garante da ética de qualquer SAI. É fácil sentirmo-nos tentados a pensar que as divulgações integrais dos dados de treino, juntamente com o código-fonte, podem conduzir a processos de tomada de decisão transparentes. Tais práticas não são suficientes. Em primeiro lugar, apenas uma pequena minoria de cidadãos seria capaz de compreender os dados divulgados e as regras algorítmicas. A transparência sem compreensibilidade apenas cumpre um objetivo muito limitado e não é satisfatória como meio de responsabilização pública. Em segundo lugar, existem limites legais a essa divulgação. A proteção do segredo comercial ou dos dados pessoais, podem constituir limites à divulgação pública (Finck, 2019; Leslie, 2019; Loi, 2021b).

Dois conceitos de transparência: técnica e do processo:

- Transparência técnica: alcançar a transparência técnica total é difícil, e possivelmente até impossível, para certos tipos de SIA. Dadas as limitações tecnológicas, talvez seja irrealista esperar que, em todas as circunstâncias, os sistemas de apoio à decisão automática sejam capazes de gerar explicações completas para as previsões que fazem. A exigência de um algoritmo gerar uma explicação dependerá muito do contexto em que o algoritmo é utilizado e do tipo de decisão para a qual contribui;
- Transparência processual: A fim de permitir a auditabilidade e a responsabilização, o que é necessário é conceber hardware, software e processos de uma forma que permita a supervisão, revisão e escrutínio do início ao fim (end-to-end).

A transparência técnica, por si só, não é suficiente para garantir ao utilizador a inteligibilidade do processo decisório. Se um cidadão desejar contestar uma decisão que foi tomada com a ajuda de um algoritmo, deve-lhe ser proporcionada a possibilidade de examinar a previsão do algoritmo e receber um resumo inteligível dos fatores que o modelo levou em conta e como estes influenciaram a previsão (Guidotti, 2018; Oswald, 2018b; Mitrou, 2021).

2.1.1. Abordagens para garantir transparência

Embora o termo "caixa preta" seja usado para se referir aos algoritmos ML/DL, os métodos de aprendizagem automática variam no nível de transparência que são capazes de fornecer. Existem formas de conhecer e até de intervir no processo de decisão e superar o efeito "caixa preta". Desenvolvimentos recentes em modelos híbridos ou colaborativos de decisão, algoritmos interativos, explicativos, causais, contra argumentativos e outros, permitem desconstruir retroativamente o processo de tomada de decisão e examinar a estrutura de cada "árvore individual de decisão", para entender como foi tomada essa decisão. Além disso, é possível visualizar para cada variável a influência ponderada que tem na previsão do modelo e, portanto, o impacto esperado sobre o resultado, conseguindo-se assim, um grau razoável de transparência, explicabilidade e auditabilidade[6].

Muito útil para este fim pode ser a investigação realizada na área da IA explicável (IAexp) (Du et al. 2020), que desenvolveu dois tipos de técnicas para a transparência de processo: explicabilidade intrínseca e explicabilidade *à posteriori* (post-hoc analysis). A explicabilidade intrínseca é alcançada através da construção de modelos autoexplicativos, incorporados diretamente nos SIA. Pelo contrário, a explicabilidade post-hoc baseia-se na construção de um segundo modelo para fornecer explicações para as decisões fornecidas pelo modelo primário (Finck, 2019; Mitrou, 2021).

A explicabilidade do SIA pode ser uma forma adequada de promover uma transparência compreensível e acionável. No cap. III teremos a oportunidade de voltar a esta questão.

Em síntese, os SIA também pode ter efeitos benéficos para a transparência. A questão não é a forma como está a ser usada, mas sim e mais relevante, a forma como foi desenhada: estes sistemas podem impactar positivamente a transparência, através da forma como são desenhados, assegurando um reforço do Estado de Direito, com uma aplicação mais consistente e coerente das regras e procedimentos formais, do que a oferecida pelos funcionários públicos. Os algoritmos podem ser úteis para superar os efeitos nocivos do viés cognitivo humano, evitar a discriminação e produzir melhores resultados. Tudo isto pode melhorar a segurança jurídica, eliminar distorções e assegurar

[6] Hildebrandt, 2011; Oswald, 2018b; Wachter, 2018; Finck, 2019; Pearl, 2019.

que todos os fatores relevantes são tidos em conta (Le Suer, 2016; Sunstein, 2018; Finck, 2019).

Embora seja reconhecido que um nível idêntico de transparência para todos os sistemas nem sempre é viável, ele deve ser o mais elevado possível dentro das restrições impostas por compromissos com outros objetivos e a regulamentação nesta matéria deve estabelecer regras e padrões claros e exigentes. A título de exemplo, os governos alemão e do reino unido destacaram que, quando a IA é usada em processos administrativos, deve ser compreensível, de forma completa e satisfatória, para o cidadão porque é que determinado órgão administrativo chegou à decisão que tomou (House of Lords, 2018; van Acken, 2019; Finck, 2019).

2.2. Explicabilidade & Direito à Explicação

A necessidade de explicação e de fundamentação, são um dos mais importantes requisitos funcionais – ao lado da transparência e da accountability, para uma IA de confiança, tendo conduzido, nas últimas décadas, ao surgimento de dois conceitos e respetivas metodologias de investigação, diferentes, mas conectados: interpretabilidade e explicabilidade.

Interpretabilidade é o processo de fornecimento de informações que representa tanto o raciocínio do algoritmo de ML, quanto a representação interna dos dados, num formato apenas interpretável por um especialista em ML. Responde ao "como" e é o primeiro passo para se chegar à explicabilidade. A explicabilidade consiste em fornecer informações num formato autossuficiente e acessível a um utilizador normal. Responde ao "porquê" das decisões. Os seus benefícios são de uma relevância extrema, quer na promoção da confiança entre os utilizadores e os SIA, como na redução da litigância e na garantia da segurança jurídica[7] [8]. Felizmente, estão já em desenvolvimento novas formas para ultrapassar estes constrangimentos (cfr. veremos no cap. III).

[7] "a explicabilidade é o núcleo central da relação de evolução entre humanos e máquinas inteligentes", Knight, W. 2017.

[8] Doshi-Velez, 2017; Oswald, 2018b; Wachter, 2018; Gilpin, 2019.

Direito à explicação - União Europeia e o artigo 22º do RGPD

Uma das razões para os progressos recentes nesta área deve-se à entrada em vigor do RGPD na União Europeia, que contém disposições que, indiscutivelmente, conferem aos indivíduos afetados por decisões puramente algorítmicas o «direito a uma explicação», cfr. artigo 22º, apesar de relativamente vagas conferem uma série de limitações, podendo condicionar as decisões administrativas inteiramente automatizadas, inviabilizando a sua operacionalidade, se as mesmas não forem explicadas ao sujeito visado, em detalhe e sob uma forma inteligível. Deste modo, fica excluída a decisão que se funda em ML, desde que o seu funcionamento não seja fundamentadamente explicável, se o mesmo é o fundamento exclusivo da decisão (House of Lords, 2018; Wachter, 2016; Bensamoun, 2022).

Em síntese, a explicabilidade dos SIA é uma novidade, ainda por explorar no âmbito judicial e administrativo, pelo que melhorar a sua performance proporcionará mais transparência sobre o modo como as decisões são tomadas, permitindo que todos os envolvidos acompanhem e percebam o processo, garantindo maior clareza na responsabilização das autoridades públicas. Explicar as razões por detrás das previsões, pode transformar um modelo ou previsão menos confiável em confiável, sabendo-se que a confiança é crucial para uma interação homem-máquina eficaz (Ribeiro, 2016; Cobbe, 2019; Pi, 2021).

Aqui chegados, deve estabelecer-se uma distinção entre as explicações sobre a forma como uma decisão foi tomada e as razões pelas quais essa decisão foi tomada, pois que as primeiras podem não cumprir, cabalmente, a obrigação de fundamentação. A questão essencial num ato administrativo, não é a descrição técnica da forma como a decisão foi tomada, mas os motivos por detrás dela - por que razão a decisão tomada foi aquela e não outra. (Edwards, 2017; Berman, 2018). A explicação pode ser usada para analisar os critérios fundamentais para uma determinada decisão (Deeks, 2019; Coglianese, 2019). Mas isso só nos leva a meio caminho, pois que a obrigação de fundamentação tem um outro patamar de exigência (Ashley, 2017).

Se um ato administrativo requer uma fundamentação jurídica substancial e complexa, então o atual nível de tecnologia não está ainda preparado, implicando a necessidade de colocar o agente administrativo no “circuito da decisão”, no controlo do

algoritmo e na satisfação da exigência legal do "direito à explicação" (Lember, 2019), como veremos mais frente no cap. III.

2.3. Accountability e Responsabilidade

A responsabilização é o princípio segundo o qual uma pessoa só pode ser responsabilizada se tiver um certo grau de controlo, no sentido de que facilitou ou causou os danos ou se estiver em posição de os prevenir ou atenuar.

No caso dos SIA quem é o responsável? O designer, o fornecedor, o utilizador, as autoridades ou o próprio sistema? A atribuição da responsabilidade complica-se pelo facto de não ser claro quem tem o controlo necessário à imputação de responsabilidade jurídica. Acresce a tudo isto, o facto de que os SIA agem de forma autónoma, aprendem e modificam a sua prestação, colocando-se a questão de saber quem controla ou prevê estes comportamentos subsequentes (Johnson, 2015; Wirtz, 2018; Gualdi, 2021).

Podemos ter uma lacuna de responsabilidade, segundo a qual os agentes públicos não podem ser responsabilizados pelo comportamento dos SIA, devido à sua falta de controlo e influência sobre os mesmos (Matthias, 2004). Porém, há quem defenda que os seres humanos são sempre responsáveis pelas consequências associadas à tecnologia, pois são eles os seus criadores (Wirtz, 2018). Estamos perante um dilema de controlo humano *versus* tecnológico, que está ainda distante de colher consenso (Anderson, 2007; Santoro, 2008; Nagenborg, 2008).

Superar este desafio é – também e sobretudo - uma questão de decisão humana, bem como de construção de consensos políticos e sociais e de sustentação doutrinal e ético-jurídica, pois que se há ou não uma lacuna de responsabilidade, ela depende mais das escolhas humanas e menos da complexidade tecnológica (Johnson, 2015; Floridi, 2018; Oswald, 2018a; Doshi-Velez, 2018).

Mesmo que os SIA não estejam sujeitos ao controlo humano direto, os funcionários e as autoridades públicas, permanecem responsáveis pelo comportamento dos SIA. O princípio da responsabilidade exige que se estabeleça uma cadeia contínua de responsabilidade para todo o desempenho do sistema (Leslie, 2019). No cap. III preconizamos uma resposta a este dilema.

2.4. Equidade & Imparcialidade

Como normalizar o enviesamento e combater o viés? Como "especificar matematicamente a equidade", de tal forma que ela possa ser auditada ou colocada como uma restrição estatística, aquando da conceção ou "treino" de um algoritmo de autoaprendizagem (Kamiran et al. 2012; Barocas, 2016; Veale, 2019).

A qualidade dos big data são um requisito incontornável para a equidade e imparcialidade das decisões suportadas em SIA - se a base de dados de partida for tendenciosa, as decisões formuladas com base nela também o serão. A qualidade da decisão depende da qualidade dos dados que foram (e continuam a ser) introduzidos no sistema (Finck, 2019). (Anastaspoulos, 2018), Tomlinson, 2020).

Algumas boas práticas nesta matéria:

a) A qualidade dos dados de entrada ou o próprio algoritmo podem ser inadequados. Por exemplo, quando alguns grupos são monitorizados mais de perto do que outros, isso pode refletir-se nos dados usados no treino dos algoritmos (cidadãos de etnia negra no policiamento preditivo nos EUA ou em sistemas de avaliação de reincidência) (Pilving, 2020).

b) Promoção e criação de conjuntos de dados mais diversificados ("dados abertos"), que reflitam de forma justa os grupos e comunidades que serão impactadas pelos SIA. Há uma variedade de grupos na sociedade, como os excluídos financeiramente e as minorias étnicas, que sofrem de "pobreza de dados", em comparação com grupos mais privilegiados da sociedade, que geram mais dados sobre si próprios (House of Lords, 2018). (Cobbe, 2019).

c) Buscar ativamente preconceitos dentro dos SIA, testando conjuntos de dados e como eles operam dentro do sistema. Se um sistema produz decisões que beneficiam ou desfavorecem consistentemente um determinado grupo, é provável que essa possibilidade exista (Cobbe, 2019).

d) Usar métodos de modelagem que levem em conta questões de igualdade (Zarsky, 2016).

e) Garantir que o desenvolvimento dos SIA é realizado por uma diversidade de disciplinas académicas, numa abordagem interdisciplinar.

Síntese do capítulo

Existe consciencialização e trabalho feito, no encontrar de soluções técnicas, para as preocupações jurídicas (éticas e outras) que a implementação dos SAI levantam, no geral e em particular no campo administrativo.

Queremos agora ver também se é possível estabelecer alguma conexão entre as duas ciências: Direito e Computação, relativas às suas formas de trabalhar a informação, analisar os problemas e decidir, não sem antes fazermos o seguinte comentário.

Ficaram por tratar outras questões, igualmente relevantes, como sejam o estatuto jurídico dos sistemas de IA, em termos de personalidade jurídica, os requisitos de índole procedimental, como sejam: a conformidade legal através do desenho do sistema; a relevância da qualidade das bases de dados; a caixa de areia regulamentar (Regulatory Sandboxes); a necessidade de auditabilidade do "technological due process", entre outos.

Capítulo II - Metodologia jurídica & inteligência artificial

Considerações introdutórias

Parece-nos fazer sentido, nesta abordagem sobre as ligações entre a IA e o Direito, não ficarmos apenas pelas considerações sobre quadro jurídico, mas incluirmos, breves notas reflexivas, sobre aquilo que é mais intrínseco às metodologias de aplicação do Direito e à forma como o raciocínio dos algoritmos evoluiu, em termos da análise da informação, da formulação do conhecimento sobre a realidade e como situa a abordagem às decisões, reflexão essa com a intenção de sabermos se existe convergência ou complementaridade.

1. Metodologia Jurídica e o processo de aplicação do Direito

1.1. Raciocínios lógicos clássicos

A metodologia jurídica como doutrina da aplicação prática do Direito, tem no raciocínio lógico-dedutivo o seu modelo ideal de aplicação do Direito, como decorrência da conceção legal-racional do Direito em que o que está em causa é a apresentação dos fundamentos da decisão, logicamente encadeados - subsumindo os dados factuais à norma e inferindo dela a consequência jurídica prevista na norma. Mas também o método indutivo assente na tese de que o Direito se estrutura num todo interrelacionado - a ideia de Direito, a norma jurídica e a decisão jurídica e realiza-se não na legislação, mas na jurisprudência[9]. Estas duas lógicas, completam-se com a abdução, que com o seu caráter aumentativo - explicativo e intuitivo -, procura alcançar o melhor resultado, utilizando o conhecimento profundo e não a melhor probabilidade matemática (Lamego, 2021, p.167).

1.2. Novas lógicas

Nos últimos anos surgiram novas lógicas, como sejam as "não-monotónicas"10, configurando que na construção da decisão jurídica não está estritamente subordinada à informação disponível, podendo ser corrigido ou revogado se nova informação for adquirida; ou a "lógica difusa" (fuzzy logic), adequada a contextos de indeterminação e

[9] Esta é uma das teses de Castanheira Neves e Arthur Kaufmann.
[10] As lógicas não-monotónicas – raciocínio corrigível ou revogável – surgiram nos anos 80 do séc. XX, no âmbito da inteligência artificial.

"autonomia valorativa", de orientação analítica, operando com base numa escala de valores aproximativos (Lamego, 2021, p. 211).

1.3. O método do caso e a jurisprudência

A jurisprudência tem como missão a revelação do sentido da aplicação do Direito e a sua atualização constante, tendo no método do caso,[11] a combinação entre a aquisição de conhecimentos jurídicos e a abertura do jurista ao "direito real", por via indutiva, através da análise e comparação das decisões judiciais (Ribeiro, 2014; Martinez, 2022, p. 166 e 331).

1.4. O sistema Jurídico, aberto, de construção indutiva, caracterizado pela incompletude, capacidade de evolução e a modificabilidade em que o Direito vai para além das normas[12] (Martinez, 2022, pp. 124; Lamego, 2021, p.129).

2. Inteligência Artificial13

2.1. Do mecanismo estático-dedutivo, ao raciocínio indutivo-dedutivo

Os algoritmos podem ser categorizados de acordo com sua natureza como determinísticos ou probabilísticos.

Algoritmos determinísticos: são convencionais e lineares por natureza, concebidos pelo homem e, para um input semelhante, produzem o mesmo output. Utilizam modelos lógico-dedutivos estáticos para representar a lei como fonte primária das aplicações de IA. Indicados para atos administrativos vinculados simples.

Algoritmos probabilísticos (de autoaprendizagem): baseados na inferência probabilística e na otimização contínuas, partindo do raciocínio indutivo para chegar ao dedutivo. Permite trabalhar abordagens que combinam a lei, com conceitos gerais, princípios e a jurisprudência, numa relação de interdependência holística. Combinam técnicas avançadas de ML e DL, que após a fase da aprendizagem usando casos reais (case-based learning) - através do treino em bases de dados estruturadas e não estruturadas -, ajustam e adaptam dinamicamente os seus próprios parâmetros de decisão,

[11] "Case method", introduzido em 1870, por Cristopher Langdell.
[12] Sistema jurídico indutivo de P. Heck e Claus-Wilhelm Canaris.
[13] Barbosa, 2021; Barthe, 2017; Brand, 2020; Buest, 2017; Coglianese, 2019; Cobbe, 2019; Côrte-Real, 2022; Duan, 2019; Hildebrandt, 2018; Kailash, 2016; Moles, 1992; Oswald, 2018a; Oswald, 2018b; Pilving, 2020; Sergot, 1990; Surden, 2014; Wiliams, 2021.

reagindo às alterações das bases de dados, procurando a otimização interna e a melhoraria da fiabilidade, criando soluções para problemas complexos, de forma a produzirem decisões precisas e fiáveis. Usam mecanismos autoaprendizagem e aprendizagem profunda, servindo-se de modelos matemáticos, que a partir de dados, constroem algoritmos, de baixo para cima, otimizando a sua prestação através da aquisição ou reorganização de novos conhecimentos, tendo por base parâmetros de ponderação.

Os algoritmos ML&DL são uma evolução da IA e dos seus primitivos mecanismos dedutivos e pré-determinados, para raciocínios de tipo indutivo, suportado no big data, nas análises matemáticas e probabilísticas e, sobretudo, em processos de autoaprendizagem, gerando o seu próprio conhecimento, através de regras de indutive learning. O algoritmo deixa de funcionar em termos meramente dedutivos, para garantir a dedução a partir da indução que ele próprio protagoniza.

2.2. Novos paradigmas

Aos algoritmos de ML&DL, faltava-lhes a vertente da relação de causalidade, tão própria do raciocínio jurídico. Esta lacuna, está já a ser combatida com soluções que a literatura da especialidade reputa de muito promissoras, com técnicas como a computação cognitiva ou a inteligência artificial causal, que abordaremos no próximo capítulo.

Em síntese, feito este "ponto de inflexão" na exposição, parece terem ficado claros os seguintes aspetos: existe uma evolução, aproximação e, até alguma simbiose, entre as metodologias e os processos de raciocínio lógico das duas ciências; boas possibilidades de aproximação das técnicas de IA ao modus operandi do Direito, no que à aplicação do Direito concerne; a relevância das bases de conhecimento – como sejam a jurisprudência -, para uma eficiente aplicação das técnicas de ML&DL e, que o direito não perdera a sua identidade e valores nucleares na abertura as ferramentas da IA.

Aqui chegados, pensamos ter feito um enquadramento suficientemente delimitador de algumas das principais questões jurídicas suscitadas por tao desafiante questão, e foquemo-nos agora no problema concreto. Afinal que possibilidades temos de aplicar as ferramentas da IA ao processo de tratamento e decisão dos atos administrativos discricionários?

Capítulo III - Modelo colaborativo de automatização jurídico-administrativo

Considerações introdutórias

Após as notas iniciais, sobre algumas das questões que devem compor a construção de um quadro jurídico, neste processo de abertura – e possível relação de parceria - do Direito Administrativo à (insidiosa) presença da IA, tendo sempre presente os perigos e cuidados a ter nessa aproximação, e considerando como conforto o "ponto de inflexão", colhido das analogias entre as duas ciências, em termos dos seus processos de raciocínio e tomada de decisão, estamos preparados para inverter a concavidade da equação e partirmos, de forma consciente e tranquila, ao encontro da questão principal: qual a viabilidade da aplicação dos SIA aos atos administrativos, e em particular aos atos discricionários?

Comecemos por uma nota introdutória à vinculação da AP:

O nosso sistema (atual) de Administração assenta no entendimento de que a eficácia do exercício da função administrativa, exige que esta seja juridicamente dotada de prerrogativas de poder público, subordinado por um lado, mas por outro, um poder dotado de autonomia em relação à lei – mas que as leis habilitam -, por insuficiência do princípio da legalidade, conferindo poderes de valoração próprios, no exercício da sua atividade. Seja através das faculdades de ação atribuídas (na liberdade de agir ou não agir); seja na liberdade de escolha de uma ou outra medida, formulando juízos de valor, dirigidos a determinar em cada caso a melhor solução para o interesse público, num poder vinculado ao correto preenchimento dos conceitos indeterminados, normas abertas ou cláusulas gerais. Procedendo nestes casos a valorações e ponderações próprios da AP, que lhe permitirão configurar as premissas em que fará assentar a justificação – fundamentação – dos seus juízos, suportados em critérios de racionalidade, exteriores às próprias normas.[14]

É neste contexto que, como refere o Mário A. de Almeida, "os procedimentos administrativos tendem a ser cada vez menos um instrumento de verificação da existência de pressupostos previamente definidos pela lei e da consequente dedução da única decisão correta a adotar em função deles, para serem um instrumento de realização de escolhas,

[14] Inspirado nos ensinamentos do Prof. Doutor Mário A. de Almeida, na sua Teoria Geral do Direito Administrativo, editora Almedina, 2021.

valorações e ponderações, em busca do compromisso entre ideias e interesses contrapostos" (Almeida, 2021, p.101).

Decisões do ponto de vista da governança pública:

Na indeclinável passagem para um "Estado Administrativo Automatizado", questionam-se premissas fundamentais no seio do direito administrativo, entre elas, duas correntes alternativas (ou complementares?) na implementação de SIA, a saber: a governança a partir de normas generalizadas , suportada nas técnicas de ML&DL, pois estas têm a capacidade de estruturar decisões a partir de contextos não estruturados ou definidos previamente (Jordan, 2015); e/ou uma governança individualizada, com "regras feitas à medida de cada cidadão" (Citron, 2008; Lessig, 1995; Kang, 2005) – sem que isso enfraqueça o poder discricionário e a equidade do caso concreto.

No futuro, poderá ser possível optar por uma ou até por ambas as alternativas, em segurança, mas isso, a acontecer, será no futuro. No momento presente, propomos um outro caminho - um modelo de balanceamento, entre a participação humana e a automatização, sobretudo para as decisões administrativas discricionárias.

A utilização de algoritmos em processos administrativos simples parece ser mais fácil de implementar. Já em processos de natureza complexa, é claramente um desafio, em que o grau de participação de humana nas decisões deve ser maior e menor a automatização (Mitrou, 2021).

Processos de natureza simples e de natureza complexa:

Inicialmente, os programas de IA – aplicados ao Direito - poderiam ser usados para suporte ao processo de tomada de decisão enquanto testam soluções. Após este período, em determinadas áreas, os processos de tomada de decisão administrativas poderiam ser gradualmente substituídos por algoritmos legais e decisões automáticas. O maior desafio é tomar decisões administrativas baseadas em regras de avaliação, cláusulas gerais e reconhecimento da discricionariedade administrativa.

Como vimos no cap. anterior, a IA pode ser usada igualmente quando queremos atingir um objetivo e não temos instruções claras de trabalho. Focando-se no objetivo a IA pode ajudar a alcançar resultados e tomadas de decisão que à partida não estão

claramente definidas (Côrte-Real, 2022, p.132). Esta possibilidade é fundamental para a criação de conhecimento e corrobora a nossa posição.

Exige também mudanças de paradigma, no sentido de passarmos de abordagens clássicas de raciocínio lógico-dedutivo para a inferência estatística e probabilística, suportadas em meta dados. Contudo são essas as formas de raciocínio de um praticante da lei, a sua forma de atuar é: "creative problem solving under conditions of uncertainty and complexity" (Okamoto, 2009).

O verdadeiro desafio para o uso dos SIA, não se encontra nos atos vinculados, mas sim nos atos discricionários, aqueles que exigem uma elevada ponderação cognitiva e subjetiva, no sentido de ser encontrada a solução ótima e justa para o caso concreto.

Propomos um modelo por níveis de automatização, que acompanhem os graus de discricionariedade dos atos administrativos, que permita navegar entre mais ou menos automatização, no sentido de evitar a restrição do poder discricionário, a insegurança jurídica e o incumprimento dos valores fundamentais.

Porque temos vários graus de atos administrativos, quanto ao seu caráter de vinculação, complexidade e outros, desde vinculados a discricionários complexos, faz sentido a criação de um modelo diferenciado de IA – mais lógico-dedutivo e abstrato no primeiro caso e mais analítico, inferencial e indutivo, no segundo caso – e de uma metodologia que combina os princípios jurídicos, a lei, a jurisprudência e a doutrina, numa relação de completividade sistémica (Moles, 1992).

Graduando os AA, em vários graus de complexidade e/ou de discricionariedade, desde fortemente vinculados a discricionários; estabelecendo níveis de automatização diferenciados, desde algoritmos de decisão automatizada, independentes do ser humano, até sistemas de apoio à decisão final ou durante o *due process*.

Para isso, parece-nos viável (e desejável) propormos um modelo matricial que combina quatro variáveis: **1. Graus de complexidade e discricionariedade dos AA; 2. Automatização e níveis de risco; 3. Sistemas diferenciados de inteligência artificial; e 4. Grau de automatização e intervenção humana,** a saber:

1. Graus de complexidade e discricionariedade dos AA

Os AA encerram diferentes níveis de vinculação e complexidade, dos simples aos complexos e dos vinculados aos discricionários, com o que podemos estabelecer – meramente para efeitos de explicação dos modelo - uma gradação por patamares: AAV; AAD simples; AAD complexos e AAD muito complexos.

2. Automatização e níveis de risco

A proposta de regulamento da União Europeia (UE) cobra aos responsáveis de sistemas de IA de "risco elevado" certas obrigações específicas, para garantir a segurança, a fiabilidade e a traçabilidade das ações, subordinadas a algoritmos inteligentes. Deste modo, os SIA devem ser enquadrados por um sistema de gestão de risco, ao longo de todo o seu ciclo de vida, por forma a serem testados e atualizados periodicamente.

No Canadá, uma diretiva consagrou em 2019 a implementação de um sistema de avaliação do impacto dos algoritmos sobre as decisões administrativas. O modelo constitui uma abordagem pelo risco, onde a proposta de regulamento da UE se parece ter inspirado (Directive sur la prise de décisions automatisée, 2019) e que seguimos de perto, como boa prática na estruturação da avaliação do risco.

Refere o seu art.º 4º, nº 1 que "*O objetivo é garantir que os sistemas automatizados de tomada de decisão sejam implantados de maneira a reduzir o risco para os canadenses e as instituições federais e que possam resultar em tomadas de decisão mais eficientes, precisas e compatíveis com a lei canadense*." O anexo A propõe uma estrutura de avaliação do impacto dos algoritmos que "ajude as instituições a melhor compreenderem e a mitigarem os riscos associados aos sistemas automatizados de tomada de decisão, fornecendo os requisitos apropriados de governança, monitorização, estabelecendo tipos de relatórios e de auditorias que melhor correspondam ao tipo de aplicativo projetado." E para essa avaliação são propostos níveis de avaliação de impacto em 4 patamares (anexo B):

I - A decisão terá pouco ou nenhum efeito e os seus efeitos são reversíveis e breves; II - É provável que a decisão tenha um impacto moderado e os seus efeitos serão reversíveis e de curto prazo; III - A decisão provavelmente terá um alto impacto e

conduzirá a efeitos contínuos, que podem ser difíceis de anular; IV - A decisão provavelmente terá um impacto muito alto e as decisões deste nível geralmente levam a efeitos irreversíveis e permanentes.

E cada nível avalia um conjunto de critérios constantes: (a) os direitos de indivíduos ou grupos; (b) a saúde ou bem-estar de indivíduos ou comunidades; (c) os interesses económicos de indivíduos, entidades ou comunidades; (d) a sustentabilidade contínua de um determinado ecossistema.

Por fim recomenda requisitos, por nível de impacto, para a tomada de decisão administrativa (anexo C), em que, para a categoria "supervisão na tomada de decisões", a intervenção do agente público, deve: a) para os níveis I e II as decisões podem ser tomadas sem a participação humana direta; b) para os níveis III e IV as decisões não podem ser tomadas sem que haja pontos definidos de intervenção humana durante o processo decisório e a decisão final deve ser tomada por um humano.

3. Sistemas diferenciados de inteligência artificial

3.1. Para os **AA vinculados**, são bastantes os algoritmos de natureza determinística e SIA de tomada de decisão automática, mantendo os cuidados referidos no cap. I.

3.2. Quanto aos **AAD simples ou medianamente complexos**, recomendamos ferramentas de ML supervisionada, que, formulam as suas decisões via mapas de aprendizagem, produzindo um output para cada input [15]. Um algoritmo de aprendizagem supervisionada, analisa os dados de treino [16] e produz uma regra geral (uma função) que é usada para mapear novos dados. Aprendizagem aqui significa ajustar o parâmetro para reduzir a discrepância, em cada caso, entre o output alvo de treino e o atual output produzido pelo modelo [17].

3.3. **AAD complexos ou muito complexos – ferramentas de ML combinadas e de DL**

[15] Jordan, & Mitchell, 2015.
[16] O treino dos dados consiste num conjunto de exemplos com os quais o computador é treinado, Sathya, 2013.
[17] Dasgupta, 2016.

Os ML não-supervisionados são mais evoluídos que os de aprendizagem supervisionada, uma vez que é o computador quem aprende, de mote próprio, a executar tarefas específicas e a extrair uma estrutura lógica de amostras de dados (mesmo de uma amostra aparentemente insuficiente e não previamente trabalhada), superando o desempenho do ML supervisionado.[18]

Destacamos, a título ilustrativo, algumas tipologias de **ML não-supervisionados**:

a) Algoritmos orientados a objetivos (goal-oriented algorithmics), centrados na otimização de uma dada função, em que o termo "otimização" implica alcançar um determinado resultado, mesmo num contexto incerto, indefinido ou complexo. (Milner, 2017).[19]
b) O uso de técnicas de clusterização, que permitem trabalhar enormes quantidades de dados e um número elevado de variáveis, conseguindo isolar categorias de informação e conceitos. Trata-se de uma técnica relevante na análise de bases de dados de informação e fundamental para o estudo e grupagem de conceitos, do conteúdo de casos reais e de jurisprudência (Milner, (2017).

c) Aprendizagem por reforço [20] [21] é um dos avanços mais recentes em termos de IA e visa treinar o sistema para agir da melhor forma possível, o que implica encontrar um modelo que maximize os objetivos estabelecidos. Aprende a resolver problemas com elevados graus de dificuldade por tentativa e erro, sendo treinado para tomar de decisões sequenciais em cenários reais. O desafio é encontrar um equilíbrio entre a exploração de território desconhecido e a exploração do conhecimento atual, maximizando as ações "corretas" a serem tomadas nos vários cenários (recompensa cumulativa).

d) Combinação de uma abordagem atomística e uma análise holística – modelo baseado na história de casos concretos. Vários estudos apontam para que em situações mais complexas e dependentes do contexto, a abordagem mais atomística e argumentativa deva ser substituída ou combinada com uma análise

[18] Ayodele, 2010; Sathya, 2012; Jordan, 2015; Dasgupta, 2016; Sushma, 2020.
[19] Abordagem que tem sido aplicada em diversas áreas, como o direito fiscal.
[20] Este algoritmo expandiu-se nos últimos anos e encontrou aplicações em múltiplos setores, como a indústria, transportes, energia, finanças, saúde, gestão, entre outras.
[21] Yeung, 2015.

holística, baseada numa visão geral dos casos concretos. [22] [23] Esta técnica combinada parece ser muito interessante para o nosso desafio dos atos dicionários, que se materializam em ambientes não estruturados, dinâmicos e de grande variabilidade (De Boer, 2008; 2005).

e) Modelos de aprendizagem profunda (deep learning models) e a importância do contexto. Estes modelos superam as ferramentas comuns de aprendizagem automática na análise de informação e na atribuição de significado a partir da semântica de textos técnicos – conseguem extrair o significado semântico, a partir da leitura do contexto, sem necessidade de recurso a meta dados estruturados (o que facilita muito o trabalho, tornando o sistema menos pesado e oneroso) (Schmidhuber, 2015; Jordan, 2015; Huang, 2021).

4. Grau de automatização e intervenção humana

4.1. Atos administrativos vinculados

A resposta à questão: **é possível um elevado grau de automatização para os AA vinculados?** - afigura-se positiva, natural e incontornável. Sim, é possível, através de: (i) algoritmos determinísticos, convencionais e lineares, concebidos por seres humanos, em que para um input semelhante, o sistema produzirá sempre o mesmo output; (ii) e de uma aproximação normativa baseada na previsão legal (Asheley, 2017; Mendes, 2020, p. 57), em que as normas são representadas mediante a sua colocação num plano lógico-formal, permitindo uma normalização fiel ao raciocínio jurídico.

4.2. Atos administrativos discricionários complexos

E quanto aos AA discricionários complexos e muito complexos?

Através de aproximações sucessivas à representação do conhecimento normativo e à interpretação jurídico-legal, das quais destacamos duas: aproximação normativa baseada em exemplos e casos concretos precedentes em que se constrói a norma através

[22] Prakken, 2009, estudo sobre a estruturas de argumentação nas decisões médico-legais; Bex, 2007 e Braak, 2007, na área da investigação criminal.
[23] Num projeto sobre investigação criminal, foi proposto um modelo baseado na história e argumentativo e uma ferramenta de software combinados, numa tentativa de combinar estas abordagens "atomísticas" e "holísticas do raciocínio probatório jurídico, Bex, 2007; Braak, 2007.

da determinação dos casos abrangidos e excluídos pela mesma. Este método é usualmente utilizado por juristas, tanto para enumeração positiva de casos-exemplo abrangidos, como para a delimitação negativa de conceitos; e a aproximação normativa baseada nos valores jurídicos.

Recorrendo a algoritmos probabilísticos e outros ainda com maior potencial de análise, e enquadrados com o procedimento administrativo, conforme veremos mais à frente, mas tendo em atenção os cuidados e ressalvas que enumeramos de seguida.

4.2.1. Constrangimentos e limitações

(1) Experiência humana acumulada na tomada de decisões discricionárias

Embora os algoritmos estejam cada vez mais preparados para substituírem a as decisões humanas, aspetos importantes dos processos de tomada de decisão (como a discricionariedade) são difíceis de serem automatizados, perdendo qualidade, equidade, entre outras. Embora exista uma experiência acumulada considerável na compreensão da tomada de decisão humana e das suas limitações, estamos apenas a começar de compreender as falhas, limitações e limites da tomada de decisão algorítmica.[24]

Existem evidências de que a tomada de decisão humana é especial (Tversky, 1974) no que diz respeito ao uso de conhecimento e normas tácitas (Schulz, 2016). Isto permite, por exemplo, que os seres humanos percebam casos excecionais em que a aplicação de uma regra não é adequada, embora o caso se enquadre no seu âmbito de aplicação. O exercício do poder discricionário refere-se à capacidade de deliberar sobre um caso e chegar a uma decisão diferente daquela que poderia ser extraída diretamente de um conjunto de regras ou protocolos. Coloca-se a questão: os humanos podem lidar com exceções por meio da discricionariedade, e os computadores podem? (Binns, 2020). A resposta exige uma melhor compreensão da conceção e das características dos processos de tomada de decisão algorítmica. (Council of Europe, 2017).

[24] "Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications". Committee of Experts on Internet Intermediaries, Council of Europe, 2017, p. 7-9).

(2) Necessidade de estabilização dos pacotes legislativos

Os SIA só podem operar com base em informações fornecidas, usando as circunstâncias existentes para aprender e extrapolar os padrões apropriados. No entanto, quer a lei, bem como a sua interpretação, não é estável, podendo mudar ao longo do tempo (Mitrou, 2021). Alterações na lei ou na sua interpretação são uma condicionante na implementação de SIA.

(3) Informação complexa, não estruturada ou conflituante

As decisões automatizadas tornam-se mais difíceis se detalhes de casos individuais ou informações ainda não estruturadas (que não fazem parte dos dados históricos dos quais as regras são extraídas) precisam ser tidas em consideração. O uso da IA parece ser questionável quando se trata de realizar tarefas discricionárias ou quando há a necessidade de estruturação ou avaliação de informações ou quando se exige o pleno conhecimento dos fatos e da complexidade do caso em questão (Mitrou, 2021; Zalnieriute, 2019; Etscheid, 2019).

(4) Alguma rigidez na análise da informação

Os programas de computador operam com base na lógica, onde as informações de entrada são processadas por meio de algoritmos programados para chegarem a um resultado predeterminado (Perry, 2014). Tal rigidez pode ser incompatível com decisões discricionárias, que precisam levar em consideração os valores da comunidade, as características subjetivas das partes ou quaisquer outras circunstâncias que possam ser relevantes (Sourdin, 2018).

(5) Capacidade de reação e adaptação humanas, atualmente mais céleres que as tomadas por SIA

As decisões administrativas desenvolvem-se e evoluem, em parte, como resultado de decisões tomadas sobre novos casos, seja no âmbito executivo ou judicial. À medida

que novas situações são decididas, novos precedentes são criados, o que influencia decisões futuras que sejam semelhantes nos aspetos mais relevantes (Binns, 2020). Tanto as decisões feitas pelos agentes públicos, quanto os processos de tomada de decisão automática, evoluem através da consideração destes novos casos, contudo existem diferenças entre ambos os sistemas (Binns, 2020). Na aprendizagem automática, novos dados pontuais podem não ser suficientes para efetuar a mesma alteração em todos os casos semelhantes no futuro. Normalmente, cada unidade de dados no conjunto de dados de treino tem peso igual; se houver mais casos anteriores nos dados de treino que tenham uma classificação diferente, eles terão um peso maior nas decisões futuras do que o novo dado. Dessa forma, novos dados adicionam, mas não anulam o modelo globalmente. Se o novo caso contradisser os casos antigos, o algoritmo de ML tentará encontrar, de forma automática, um modelo com o melhor ajuste entre todos os casos da base de dados, e para que isso aconteça vão ser necessários muitos casos. Por isso que, tais casos influenciam as decisões humanas posteriores de forma mais célere, que influenciam a decisões tomadas por SIA. (Binns, 2020).

(6) Decisões orientadas a valores e a princípios e não (somente) por bases de dados

Complexidade adicional pode ocorrer quando um sistema automatizado precisa tomar não apenas uma decisão orientada por dados, mas também por princípios e valores. Sistemas baseados em regras podem ser incorporados a um aplicativo de tomada de decisão de IA, mas torna-se mais difícil quando a AP precisa encontrar um equilíbrio entre direitos e interesses concorrentes, uma vez que existe a dificuldade em garantir que um sistema orientado por princípios, "programado com moralidade escrava" (Wirtz, 2019), não resulte em rigidez moral ou legal em relação às circunstâncias individuais.

Decidir nestes contextos implica ponderar regras conflituantes e decidir qual deve ter precedência no caso específico ou desconsiderar uma regra após a ponderação de certos fatores contextuais da situação em questão que tornam a sua aplicação inadequada. Os julgamentos que envolvam a aplicação de valores ou princípios possivelmente conflituantes é algo que, porventura nesta fase, devemos deixar aos humanos (Binns, 2020; Citron, 2008a). A aplicação de regras e critérios uniformes em todas as decisões, pode constituir um uso indevido do poder discricionário, conduzindo a uma decisão ilegal (Cobbe, 2019; Mitrou, 2021).

(7) O caso concreto e a restrição da discricionariedade

Quando está em causa uma decisão discricionária, as circunstâncias individuais devem ser tidas em conta. Se tal não acontecer podem ocorrer questões de restrição ao poder discricionário. Nesse sentido os sistemas de aprendizagem automática podem ser inadequados para decisões em que é provável que poderes discricionários precisem ser exercidos caso a caso ou em outras situações em que as normas podem ser aplicadas de forma geral, mas onde é provável que determinadas exceções precisem ser acauteladas (Cobbe, 2019).

O problema coloca-se quando a discricionariedade foi delegada ao algoritmo, e nenhuma escolha consciente está sendo feita pelo órgão ou pessoa responsável pela decisão. Sendo mais grave quando as decisões envolverem questões de direitos humanos e, portanto, necessariamente exigirem que poderes discricionários sejam exercidos com a devida ponderação (Cobbe, 2019; Hildebrandt, 2018; Oswald, 2018a).

(8) Compreensão abrangente de todo o contexto legal

Um agente público é capaz de compreender e aplicar a lei num contexto muito mais amplo: a história da lei, a sistemática das normas, o objetivo da lei, o significado geral da justiça, mas especialmente os efeitos diretos e indiretos da decisão. Não é possível em todos os campos produzir dados quantitativos ou qualitativos suficientes para descrever todas as camadas da lei e o seu ambiente operacional. E está longe de ser possível para os sistemas inteligentes (atuais) seguir todo este entorno em tempo real (Pilving, 2020).

4.2.2. Argumentos favoráveis ao uso da inteligência artificial na discricionariedade

De forma resumida alguns desses argumentos favoráveis são: (1) Os SIA podem resultar em menos discricionariedade, mas as suas decisões são percebidas como objetivas (Mitrou, 2021); (2) Uma razão válida para a introdução destes sistemas pode ser a necessidade de redução de custos e redução de pessoal (Mitrou, 2021); (3) Está assim tão claro que os humanos podem "fazer" discricionariedade, mas as máquinas não? Se a discricionariedade for essencialmente o tratamento de exceções, tanto as máquinas

quanto os humanos podem exercê-la. A discricionariedade, numa abordagem restrita, não é mais do que a capacidade de reconhecer quando uma regra geral não deve ser aplicada a um caso particular. Numa linguagem computacional é uma questão de tratamento de exceções. Por isso, quer um sistema algorítmico seja "baseado em regras" ou estatístico, é possível projetá-lo de forma a criar exceções às regras gerais. Para um sistema especialista, trata-se simplesmente de uma questão de adicionar outra regra para modificar a mais geral. Para um modelo estatístico, trata-se de incluir exemplos de exceções, que permitem ao sistema distinguir os casos que se enquadram na regra geral e os que se enquadram na exceção. Os modelos de IA têm a capacidade de capturar essa relação de exceção, desde que a mesma seja refletida nos dados de treino (Binns, 2020); (4) Alguns estudos encontraram um impacto positivo dos algoritmos no trabalho e no poder discricionário dos funcionários públicos, pois a IA foi vista como um sistema de apoio à decisão em vez de um sistema de tomada de decisão. Portanto, o papel dos humanos e da discricionariedade precisam ser fortalecidos em vez de descartados (Criado, 2020; Mitrou, 2021); (5) A decisão robótica, assim como a humana, não podem ser consideradas infalíveis. No entanto, a automatização do processo de tomada de decisão permite obter vantagens significativas em termos de uniformidade, fiabilidade e capacidade de controlo adequado da decisão;

Posto isto, a questão essencial para a qual devemos ensaiar uma resposta, prende-se com o facto de sabermos se: os SIA podem ser usados como decisão automática completa e final, ou se devem antes ser um processo de suporte e recomendação à decisão?

Os sistemas de decisão algorítmica são frequentemente apresentados como substituindo os tomadores de decisão humanos, mas muitos sistemas não se destinam a substituir totalmente os trabalhadores (Kellogg, 2020).

Em vez disso, eles fornecem suporte à decisão (recomendações, previsões, análise de resultados), enquanto os trabalhadores mantêm a capacidade de ajustar ou anular as decisões finais sugeridas pelo sistema (Levy, 2021).

Ao nível da AP existe consenso generalizado de que pese embora os avanços tecnológicos alcançados, ainda é prematuro para certos atos ou decisões eliminar o recurso ao decisor humano e que, mais relevante do que substituir um pelo outro, é

pensar em como homens e máquinas devem interagir para construírem uma decisão em conjunto.[25]

Deste modo, para decisões administrativas complexas, é necessário manter o controlo e o poder discricionário do lado dos agentes públicos, assumindo a IA, o papel de suporte à decisão, fornecendo recomendações aos atores humanos, ou pelo menos os humanos devem ter um papel relevante "no ciclo da decisão" (Mitrou, 2021).

Em resumo, diríamos que os SIA podem ser aplicados aos AA vinculados e aos AA discricionários simples, como sistemas de tomada de decisão autónoma, mas relativamente aos AA discricionários mais complexos, pode ser arriscado confiar a decisão final aos algoritmos, devendo ser deixado espaço para a intervenção do agente público, até porque em qualquer das circunstâncias, a responsabilidade final nunca deixa de pertencer ao decisor administrativo.

Nesse sentido, devemos seguir uma outra abordagem. É o que propomos no ponto seguinte.

4.3. Proposta de um modelo de inteligência colaborativa homem-máquina

Com acabamos de referir, a tomada de decisão totalmente baseada em IA parece ser aceitável apenas nos casos em que há baixo nível de discricionariedade ou mesmo ausência de discricionariedade, por ex. quando o órgão público estiver estritamente vinculado à lei em relação às disposições e procedimentos a serem seguidos e não forem razoavelmente presumidos efeitos negativos para indivíduos e/ou grupos.

O algoritmo não pode escapar à administração, que deve controlar a conceção, o funcionamento e as suas evoluções, sob pena de incompetência negativa[26]. Lembremos que a autorização dada à AP de tomar uma decisão administrativa individual exclusivamente suportada por um algoritmo, não corresponde a um abandono das suas competências, uma vez que as regras e os critérios seguidos pelos algoritmos são parametrizados pela administração (Bensamoun, 2022, p. 505). As pessoas encarregues deste controlo humano podem testar e corrigir os SIA, reinterpretar, ignorar, anular ou

[25] Coglianese, 2017; Arteaga, 2020; Guay, 2018; Kellogg, 2020; Levy, 2021; Veale, 2019; Vogel, 2020; Milner, 2017; Vogel, 2019; Yeung 2017; Finck, 2019.
[26] O controlo dos SIA pelo homem está no centro das reflexões sobre ética dos algoritmos. O Homem deve poder "manter a sua mão" ("garder la main" ou "human-in-the-loop").

reverter o resultado do sistema de IA e ainda interromper o seu funcionamento, cfr. art.º 14º, nº 4 alíneas a) a e) (Comissão Europeia, 21 de abril 2021, COM (2021) 206 final, Regulamento Inteligência Artificial).

Caso o nível de discricionariedade seja elevado, devido aos dados em causa e às circunstâncias específicas que devem ser tidas em consideração, os SIA podem servir de suporte à tomada de decisão humana, por exemplo, compilando informações automaticamente, processando informações ou dando sugestões para avaliação (Etscheid, 2019; Mitrou, 2021)[27].

Uso da IA como um meio para o incremento e a ampliação da atividade humana e não como um meio para a sua automatização (isto é, substituição do homem pelos algoritmos). A IA vista como um sistema de suporte à tomada de decisão e não como um sistema de tomada de decisão; evolução de um sistema de decisão automatizada, para um sistema de suporte à decisão, com supervisão humana.[28]

E se, em vez de colocarmos a questão óbvia: e se as tarefas atualmente executadas por humanos fossem, em breve, realizadas de forma mais económica e rápida por máquinas, formulássemos uma outra questão: Que novos desafios poderiam ser alcançados se tivéssemos sistemas que nos ajudassem a pensar melhor e a tomarmos melhores decisões? Em vez de pensarmos na automatização como uma equação de mera substituição do homem pela máquina, mas sim como uma colaboração homem-máquina? (Davenport, 2015).

Davenport e Kirby desenvolveram uma abordagem que faz a integração da atividade humana com os sistemas inteligentes, enfatizando o "aumento em vez da automatização" (augmentation, more so than automation) (Davenport, 2015).

Na "automatização" a intenção é eliminar o ser humano assim que as tarefas estejam codificadas, informatizadas e automatizadas; na "aumentação" começa-se por identificar o que os humanos e os SIA podem fazer individualmente; depois é feita uma

[27] Por exemplo, na política social, a IA tem sido usada para apoiar a previsão de jovens de alto risco para direcionar intervenções (Sun, 2019) ou para permitir previsões mais precisas para detetar creches para crianças ou para referenciar famílias que podem carecer de mais inspeção (Sistema Kind en Gezin desenvolvido pela Agência Flamenga para a Criança e a Família), Mitrou, 2021.

[28] Adaptamos e seguimos de perto um modelo proposto para o setor privado, que faz o escalonamento da implementação dos sistemas de IA em três níveis: situações e contextos que podem ser totalmente automatizados; contextos que não podem ser automatizados e os que estão numa espécie de "missing middle" e que recomendam um modelo de inteligência colaborativa homem-máquina.

análise de como esse trabalho poderia ser aprofundado e incrementado pela colaboração entre os dois, sempre com o propósito de possibilitar que as pessoas passem a fazer um trabalho mais valioso, criativo e intelectualmente estimulante. A IA pode aumentar as nossas capacidades analíticas e de tomada de decisão, fornecendo a informação certa no momento certo (Daugherty, 2018b). "Aumento é mais do que automatização", é uma forma de mitigar a não automatização total, através da inteligência aumentada e da simbiose homem-máquina (Miller, 2018).

Em 1960, Licklider publicou um artigo intitulado "Man-Computer Symbiosis" onde formulou o objetivo de conseguir que pessoas e computadores cooperem na tomada de decisões e no controle de situações complexas sem dependência inflexível de programas predeterminados. Já nessa época, ele observava que "a parceria simbiótica realizará operações intelectuais com muito mais eficácia do que o homem sozinho consegue realizar" [29]

Englebart é reconhecido como o criador do conceito de "Inteligência Aumentada" (Engelbart, 1962). A "inteligência aumentada" é a expansão das faculdades cognitivas humanas decorrentes do uso de ferramentas inteligentes[30].

Mais recentemente, dois consultores da Accenture, Daugherty e Wilson, desenvolveram um trabalho muito interessante ao criarem uma estrutura sobre modelos colaborativos homem-máquina, no seu livro, "Humans + Machine: Reimagining Work in the Age of AI", publicado em março de 2018. Nele afirmam que "a verdade é que as máquinas não estão a dominar o mundo, nem a eliminar a necessidade de humanos no local de trabalho. Nesta era atual de transformação dos processos de trabalho, os sistemas de IA não nos estão a substituir de forma massiva; pelo contrário, eles estão a ampliar as nossas competências e a colaborar connosco para alcançarmos ganhos de produtividade, que antes não eram possíveis." (Daugherty, 2018a).[31]

[29] Os conceitos de inteligência aumentada e simbiose homem-máquina remontam às atividades dos pioneiros da computação nos Estados Unidos, Licklider e Englebart na década de 1950, e aos seus respetivos escritos publicados no início dos anos 60.

[30] O conceito de "inteligência aumentada" como forma de complementar e ampliar as capacidades humanas de cognição e colaboração, por meio de novos tipos de simbiose entre pessoas e máquinas cada vez mais inteligentes

[31] Outros autores falam numa simbiose homem-máquina, e fazem um apelo no sentido que os humanos e as máquinas precisam de trabalhar de forma recíproca, para aumentarem as capacidades uns dos outros. Duan, 2019; Miller, 2018; Miller, 2017; Davenport, 2018; Jarrahi, 2018.

Segundo estes autores o esforço contínuo de transformação proporcionado pela IA criou um espaço enorme, dinâmico e diversificado no qual homens e máquinas colaboram para alcançar ganhos significativos de qualidade, eficiência e desempenho dos processos e dos negócios, espaço esse que deram o nome de "missing middle" (Daugherty, 2018a) [32].

Dimensão 2: Funções que exigem elevado grau de liderança, relacionamento, processos abertos e/ou menos estruturados e julgamento holísticos, permanecerão como atividades exclusivamente humanas. Dimensão 1: Do outro lado do espectro, trabalhos que exigem elevados graus de execução de transações, interação repetível, previsão baseadas em modelos e as adaptações que podem ser especificadas por modelos ou regras ou derivadas de dados, tornar-se-ão cada vez mais atividades de automatização exclusiva.

Dimensão do meio ou esquecida: nas atividades que se situam entre a 1 e a 2 - o meio que falta ("the missing middle") encontramos uma crescente gama de atividades de trabalho, que são melhor executadas através de um esforço conjunto de inteligência colaborativa homem-máquina [33] .

Em resumo:

Embora a IA tenha o potencial de substituir muitas das funções administrativas, é mais provável que os avanços tecnológicos sejam no sentido do "apoio" do que da substituição (Sourdin, 2018). A estratégia parece ser a da implementação de sistemas de IA que, complementem o trabalho humano atual, permitindo maior eficiência, aumentando o talento e a performance humana e dos processos de trabalho, em vez da substituição total de humanos, através de uma automatização tradicional que num futuro mediato conduza à estagnação, tudo isto se traduzindo, não na questão de se saber se a Administração Pública está a implementar a IA, mas como ela o está a fazer! (Surden, 2014; Sourdin, 2018; Daugherty, 2018a).

[32] A esse espaço os autores deram o nome de "missing middle" – processos que estão num espaço do meio, esquecido ou não aproveitado, em que "preencher o meio que falta", significa trabalho em conjunto através da definição de novos tipos de papéis e parcerias colaborativas.

[33] Através dessa inteligência colaborativa, os seres humanos e a IA melhoram ativamente os seus pontos fortes e completares: a liderança, o trabalho em equipa, a criatividade e as competências sociais dos primeiros, e a velocidade, a escalabilidade e as capacidades analíticas dos segundos, Daugherty, & Wilson, 2018b.

Conceber um modelo jurídico-administrativo operativo, será também uma forma de seguir aquilo que nos mais variados contextos, empresarias e outros, se procura adotar, isto é, construir sistemas que operam num continuum entre a automatização total e a robótica colaborativa, na era da indústria 5.0 (que no caso em apreço se pode traduzir no suporte a decisão administrativa). Além disso, criar sistemas de IA com uma mentalidade de aumento (aprimoramento da simbiose homem-máquina) em detrimento duma postura de automatização (de mera substituição do homem pela máquina) faz toda a diferença para a motivação e realização dos profissionais, assim como para a segurança nas decisões (Miller, 2018; Economist Intelligence Unit 2017, sobre o uso de IA na indústria; Davenport, 2015). [34]

O uso da IA para "aumentar", em vez de substituir o trabalho dos agentes publicos, pode ser uma ferramenta eficaz para melhorar a qualidade dos serviços administrativos, operando dentro dos parâmetros e princípios do direito administrativo[35] (Yamane, 2020).

4.4. Soluções algorítmicas, já hoje disponíveis e que sustentam o modelo proposto

Existem atualmente um conjunto de novos algoritmos, dentro do ML ou até superiores a estes, que suportam o modelo proposto de interação colaborativa ou simbiose entre o utilizador dos sistemas e os algoritmos e auguram excelentes perspetivas no processo de dotação dos procedimentos administrativos e os AA discricionários de ferramentas apropriadas. Segue um apertado resumo:

[34] A IA pode desempenhar múltiplos papéis na tomada de decisões, mas a IA será maioritariamente aceite pelos decisores humanos como uma ferramenta de apoio à decisão e não como a automatização da tomada de decisão para os substituir.

[35] Exemplo de um projeto que combina a supervisão humana e automatização, o projeto Taxman é uma experiência na aplicação de técnicas de inteligência artificial ao estudo do raciocínio jurídico e da argumentação jurídica, utilizando o direito fiscal das sociedades (Cook, 1981).

(a) O Homem no circuito – (human in-the-loop) [36]

Um "humano-no-circuito"[37] é uma forma de inteligência colaborativa homem-máquina, em que o utilizador final, participa no processo de tomada de decisão algorítmica (Monarch, 2021).

Este processo está relacionado com as preocupações de que, à medida que os algoritmos se tornam mais complexos e são usados para tomar mais decisões, há uma necessidade de os utilizadores finais especialistas estarem envolvidos, a fim de garantirem que essas decisões são eticamente sólidas. O "humano-no-circuito" estabelece métodos e processos para que humanos e algoritmos trabalhem em conjunto, de forma a otimizarem todo o processo de aprendizagem automática e a tornarem o resultado mais eficaz (Monarch, 2021).

Atualmente, a maioria dos sistemas de aprendizagem automática implementados aprendem com o feedback humano, que se torna relevante na melhoria das previsões e na verificação da sua precisão, sugerindo correções e aperfeiçoamentos. Torna o processo mais transparente, uma vez que, ao incorporar julgamentos humanos em cada etapa do processo, as pessoas podem melhor compreenderem porque certas decisões foram tomadas. Ajuda a construir confiança entre humanos e sistemas automatizados e torna o sistema muito mais rápido e eficaz, comparado com um sistema totalmente automatizado (Oswald, 2018a).

Por outro lado, algoritmos implementados no setor público poderão estar limitados na alimentação de dados, por razões legais, técnicas ou políticas, sendo necessário recorrer ao agente público, pois só ele tem o conhecimento dos fatores que "não podem ser representados por entradas no algoritmo", incluindo o conhecimento processual e tácito adquirido pela experiência e pelas boas práticas (Grove, 1996; Lee, 2018; Oswald, 2018a).

Daí que seja prudente e recomendado que o agente público, nos tempos mais próximos e para as decisões discricionárias, esteja comprometido durante todo processo (human-in-the-loop) e não no final do mesmo, como mero "carimbador de selos"

[36] Human-in-the-loop, é uma metodologia de desenvolvimento de SIA centrada no utilizador final, isto é, o centro de todo o processo está suportado e centrado na experiência e competência do utilizador (Monarch, 2021)

[37] Um "humano-no-circuito" refere-se a um SIA que depende de feedback ou inputs do ser humano. Ao contrário, um SIA que não depende dos inputs humanos, é designado de "humano-fora-do-circuito".

(Oswald, 2018a). Este processo permite também dar resposta às recomendações da EU, sobre o uso de sistemas de IA suportados no envolvimento e intervenção humana.

(b) ML interativa (MLi) e de "caixa transparente" (glass-box) [38]

Complementar da abordagem "o homem-no-circuito", está uma outra que promove a interação homem-máquina, numa perspetiva de "caixa transparente" (glass-box) para resolver os problemas da "black box" e da análise e tomadas de decisão com elevados níveis de incerteza, colocando o decisor humano no circuito da decisão e não apenas no final da mesma.[39]

Consiste num processo de aprendizagem automática centrada no utilizador. Não se trata de colocar o humano no controlo meramente físico do procedimento de tomada de decisão ou no final do mesmo (como o "homem do carimbo"), mas usar as suas capacidades cognitivas para compensar as abordagens automáticas.

O MLi significa a integração de um ser humano no "circuito algorítmico", ou seja, passar de uma abordagem de "caixa preta" para uma "caixa de vidro", sendo constituído por algoritmos que interagem com os utilizadores especializados na matéria em causa, otimizando a sua performance de aprendizagem e gerando confiança no sistema. Consequentemente, a abordagem MLi[40] é eficaz na resolução de problemas, onde não existem grandes bases de dados, ou quando é necessário lidar com dados complexos ou eventos raros, onde a abordagem de aprendizagem automática pode sofrer de viés ou resultados insatisfatórios, por insuficiência de uma adequada amostra de dados no seu processo de treino.

---

[38] Akgl, 2011; Amershi, 2014; Gigerenzer, & Gaissmaier, 2011; Holzinger, 2016; Holzinger, 2017; Kieseberg, 2016a; Kieseberg, 2016b; Robert, 2016; Schirner, 2013; Shyu, 1999; Wilson, 2015.

[39] Teng Lee desenvolveu o "Transparent Boosting Tree", combina as vantagens do ML e inputs humanos no processo de tomada de decisão final. Permite visualizar o modelo de tomada de decisão, levando em consideração o feedback de cada utilizador, incorporando-o sucessivamente no processo de aprendizagem e de melhoria da performance do algoritmo Lee, 2016.

[40] O uso da metodologia iML na área da saúde tem crescido exponencialmente. A inclusão de um "doctor-into-the-loop", pode desempenhar um papel significativo no apoio à resolução de problemas difíceis, em combinação com técnicas de Crowdsourcing (modelos que utilizam o conhecimento e a aprendizagem coletivos para a resolução de problemas ou desenvolvimento de uma solução), Robert, 2016.

c) A Inteligência Artificial explicável (IAexp)[41]

Para resolver os problemas da crescente opacidade, e consequente falta de transparência, os especialistas informáticos, desenvolveram um modelo, de explicabilidade intrínseca, designado de "IA explicável", em duas abordagens: (a) uma focada em descrever como as caixas pretas funcionam (open the black boxes) – "abordagem técnicista"; e (b) outra centrada em explicar ao utilizador final as próprias decisões, utilizando métodos explicativos extrínseco, [42] sem se preocupar com o funcionamento dos sistemas de decisão, chamada de "abordagem exógena". Esta abordagem não requer que o utilizador final entenda a lógica interna do SIA, para fazer uso dele, fornecendo informações que são facilmente entendíveis e úteis para compreender as razões de uma decisão.

As abordagens exógenas podem ser de dois tipos, centradas no sistema ou no sujeito. A centrada no modelo, conhecida como "explicabilidade global", envolve uma explicação exaustiva das várias componentes da criação, teste e forma de funcionar do modelo não (demasiado) técnicas, em vez do seu desempenho em um caso específico, fornecendo evidências sobre se as decisões são tomadas de forma processualmente regular. A centrada no sujeito/utilizador, também conhecida como "explicabilidade local", explica como o sistema tomou uma decisão específica num caso particular, fornece informações sobre as características dos indivíduos que foram objeto de decisões semelhantes, permitindo ao utilizador a navegar e simular o seu caso concreto, desafiando-o a encontrar e valorizar as várias alternativas.

Referir que a IAexp tem vindo a colher grande aceitabilidade nas mais variadas áreas de atividade e a dar uma resposta positiva à necessidade de transparência e de modelos explicativos acessíveis. Precisando, contudo, melhorar a relação entre explicação e grau de precisão e eficiência do algoritmo.

[41] Adadi, 2018; Edwards, 2017; Deeks, 2019; Friedler, 2018; Guidotti, 2018; Kroll, 2017; Pi, 2021; Rudin, 2019; Sarkar, 2018; Selbst, 2018; Wachter, 2018.
[42] Como se de um sistema pedagógico explicativo se tratasse, Edwards, 2017.

d) A IA causal (IAc)[43]

A IAc identifica e usa relações precisas de causa-efeito, que vão mais além dos modelos baseados em correlações, e aposta em sistemas de tomada de decisão mais eficazes, colaborativos e explicáveis. Ao identificar a causa-efeito, a IAc fornece um processo de análise (mais) semelhante ao humano, incutindo mais confiança nos utilizadores deste tipo de sistema face aos modelos convencionais de ML. Consegue elaborar "mapas causais", que modelam o processo de geração de conhecimento, em vez de realizarem simples associações entre variáveis - conseguindo uma simbiose eficaz entre o conhecimento dos especialistas e a abordagem baseada em dados (chamada "inteligência de decisão" ou modelo causal estrutural, em que as relações entre as variáveis representam efeitos causais).[44]Uma análise da causalidade pode ser usada para complementar as decisões humanas, em situações em que a compreensão das causas por trás de um resultado é necessária.

e) Computação cognitiva (CpCg)[45]

A CpCg é uma nova era da IA, iniciada por volta de 2011. Sistemas particularmente eficazes na manipulação de dados não estruturados,[46] estando entre as melhores estratégias, em termos de eficiência e qualidade daa soluções, para problemas de elevada dimensão e complexidade[47]. São uma mistura de ciência da computação com ciência cognitiva, ambicionando fazer coisas que apenas os humanos são capazes de fazer, contudo, fazendo-o de forma mais rápida e com maior precisão. Prevê-se que no futuro venham a afetar todos os setores de atividade, privada ou pública, e onde quer que hajam dados e um problema a ser resolvido, a CpCg pode desempenhar um papel relevante.

[43] Sgaier, 2020; Sharma, 2022; Shekhar, 2022.
[44] Extraído de CausaLens.
[45] Holzinger, 2017; Mehta, 2019.
[46] Um estudo de 2015 do International Data Group estima que cerca de 90% dos dados gerados atualmente não são estruturados.
[47] Alguns dos algoritmos são os desenvolvidos no âmbito da "Inteligência de Enxame", chamados de "otimização da colónia de formigas", com capacidade exponencial e dinâmica de adaptação a novos problemas em tempo real, Holzinger, 2017.

Em síntese

Na busca de maior eficiência, usufruindo do incremento das decisões administrativas automatizadas, o legislador teria de proceder à substituição dos princípios discricionários por disposições a "preto e branco", simplificando e tornando a lei mais objetiva e determinante, permitindo assim o uso de decisões automatizadas administrativas (e não só), mais lineares e padronizadas (Perry, 2017; Roth, 2016).

Tais emendas podem resultar em decisões injustas ou arbitrárias devido à falta de justiça e discricionariedade individualizadas e à falta de nuances na lei, que permitam a proporcionalidade, a gradação ou a distinção fina e sensível de cada caso (Sourdin, 2018).

Tradicionalmente, verifica-se o inverso, isto é, o tomador de decisões públicas tem a necessidade e a legitimidade de uma certa liberdade decisória, sobre a forma como cumprir a legislação e as políticas públicas. Desta forma, o contexto e as circunstâncias específicas devem ser levados em consideração ao tomar essas decisões, possibilitando assim soluções mais aceitáveis, mas, ao mesmo tempo, a discricionariedade pode resultar em tratar os indivíduos de maneira diferente (Mitrou, 2021).

Poder discricionário: Estes problemas são exacerbados por decisões discricionárias em que a lei não prescreve instruções claras. Contudo, o poder discricionário não dá às autoridades o direito de tomar decisões arbitrárias. As decisões discricionárias devem igualmente obedecer aos princípios gerais da justiça e considerar o objetivo da lei e todos os factos relevantes específicos de cada caso. Um algoritmo pode não ser adequado para tomar decisões discricionárias, porque as circunstâncias são imprevisíveis e muitos casos são atípicos, conduzindo a decisões no limite das normas (Pilving, 2020).

Atualmente existe ainda pouca evidência científica para conhecermos como é que a introdução de uma ferramenta de suporte algorítmico pode afetar os elementos discricionários da tomada de decisão (Oswald, 2018a).

No entanto, o direito administrativo está gradualmente evoluindo sua visão sobre as políticas, com aceitação crescente de que a política aplicada de forma consistente (com exceções apropriadas quando necessário para acomodar casos incomuns) pode

fornecer benefícios para boa governança, consistência e previsibilidade (Le Suer, 2016; Cobbe, 2019).

Embora a capacidade dos algoritmos de última geração, com aprendizagem automática, seja capaz de aproveitar os parâmetros discricionários dinâmicos, e conseguirem trabalhar em breve dentro das linhas complexas dos princípios de valorização e dos limites discricionários. Isto não parece realista no momento atual. As decisões deste tipo são demasiado únicas e especificas, não cabendo – em tempo útil - dentro da geração de grandes bases de dados, para que a aprendizagem automática seja capaz de os moldar. Delegar plenamente uma decisão discricionária ou baseada em juízos complexas a um algoritmo constituiria, em nossa opinião, uma violação grosseira da discricionariedade. Um algoritmo pode, no entanto, ser implementado como uma ajuda (Pilving, 2020).

Nestas notas sobre um modelo preditivo no âmbito dos atos administrativos, avulta ainda a questão da responsabilidade pela decisão, e relativamente à qual se coloca a questão, tudo ponderado, de afinal não haver diferença entre fundamentação, mediante a articulação das relações de causalidade, e a explicação das correlações operadas pelo algoritmo dos resultados obtidos: ambas visam racionalizar o processo de tomada de decisão. Se aceitarmos que relativamente às decisões dos SIA, nelas podem vir a entrar ou já entram intuições, e que o efeito black box pode ser ultrapassado, serão ambas boas decisões. Mais. Não basta dizer que as decisões humanas implicam uma decisão ética e os algoritmos não compreendem a ética: "não há motivo para pensarmos que os algoritmos não conseguirão ter um desempenho melhor do que o ser humano médio, até nesta área". Os algoritmos podem seguir diretrizes éticas, e se os programarmos para ignorar certas variáveis, podemos ter a certeza de que o computador ignorará esses fatores, porque não têm subconsciente (Harari, 2018; Rodrigues, 2020, p. 52).

Resta saber se, ao entregar a decisão à máquina, o homem não lhe transfere, desta forma, a responsabilidade da decisão que lhe deve caber a ele, enquanto agente moral, responsável pelas suas decisões em sociedade, quando estão em causa certos valores humanos. A questão a que temos de responder é: queremos produzir um artefacto preditivo para ser um agente moral? Ou queremos ser nós, enquanto criadores de produtos inteligentes, bem desenhados e transparentes, responsáveis por eles? A questão é vamos entregar todas ou uma grande parte das decisões a sistemas inteligentes porque eles

decidem melhor que os humanos? Ou estaremos a falar de uma espécie de "genocídio humano" (Greco; 2020) (nota: até podemos construir sistemas inteligentes que tomam ações morais, mas a questão reside em saber quem seria responsável por essas ações. Um agente que tome uma ação moral, não é necessariamente o agente moral responsável por essa ação. Será que a sociedade, como a conhecemos, consegue sobreviver se a responsabilidade for delegada a entidades que podem ser especificadas, construídas, compradas e vendidas? Que benefícios podemos colher com esta transferência de responsabilidade, para nós ou para a sociedade? (Bryson, 2018; Rodrigues, 2020, p. 52).

É inquestionável que está dentro da capacidade da nossa sociedade definir robots e algoritmos como agentes morais, mas, mesmo que seja tecnicamente possível, nada disto faz com que seja necessário ou que devamos fazê-lo (O facto de algo poder ser feito não quer dizer que deva ser feito). Tanto os nossos sistemas éticos, quanto os artefactos de tomada de decisão são passiveis de design humano, contudo tornar os sistemas de IA morais é uma ação intencional e evitável (Harari, 2018; Rodrigues, 2020, p. 52).

Ainda a propósito da qualificação jurídica dos SIA públicos:

No direito da UE é privilegiada uma abordagem pelo risco: a maior parte dos SIA públicos entram na categoria de "tratamento de alto risco", sendo submetidos a exigências particulares. A proposta de regulamento europeu para a IA contém uma diferenciação de regime jurídico dos SIA, em função dos riscos que podem representar para a sociedade, em termos de segurança e de atentado às liberdades fundamentais. Um tratamento de "alto risco" será desta forma submetido a exigências reforçadas. A identificação deste tipo de tratamento não é totalmente evidente, mas parece que a maior parte dos SIA públicos, em termos das suas finalidades, relevam desta categoria, tal como previsto na proposta de regulamento (parágrafo 6 e anexo III).

No direito interno dos países da EU, existe uma tendência para considerar os SIA públicos como dispositivos de ajuda/apoio à decisão. O algoritmo não é a decisão, mas uma ajuda à decisão.

Deste modo os SAI não são mais do que uma ajuda à tomada de decisão, mesmo na hipótese em que a decisão repousa exclusivamente sobre o algoritmo, uma vez que é o ser humano quem faz a parametrização do algoritmo (e uma vez que não sabemos em que medida o agente público está efetivamente capaz de se emancipar da solução proposta

pelo algoritmo). Uma tal qualificação jurídica, deve ser considerada à luz dos seus efeitos legais, nomeadamente contenciosos (Bensamoun, 2022, p. 517)

Porque as decisões algorítmicas têm um impacto determinante sobre a vida das pessoas e porque, pelo seu carácter unilateral e executório de pleno direito, são uma manifestação da expressão da força pública, um regime específico começa a tomar forma.

E, a automatização pode resultar em: (i) decisões totalmente automatizadas; (ii) modelo de inteligência colaborativa homem-máquina; (iii) ou a IA pode ser usada apenas como um sistema de recomendação em que o decisor público tem a liberdade e o poder de desviar-se da decisão sugerida.

Apesar da previsibilidade e da responsabilidade das decisões poderem ser diferentes para cada uma destas situações, em última instância a AP e os seus agentes, sempre permanecerão responsáveis. Deste modo, há uma necessidade intrínseca de supervisão humana e os tomadores de decisão devem ter a autoridade suficiente para controlar o sistema e lidar com os resultados indesejados (Mitrou, 2021).

Se os homens erram, também as máquinas com IA podem estar expostas ao erro, ou a falhas na sua tecnologia, com consequências imprevisíveis e gravíssimas. Isto significa que mesmo utilizando a IA é sempre necessária a supervisão humana, tanto mais que não podem ser colocados em risco os direitos humanos e a sociedade em que vivemos, para além de ter de existir sempre a imprescindível segurança jurídica (Santos, 2022 a) p. 21).

Este direito em construção evolui sobre uma linha de fronteira entre a possibilidade de poderem ser tomadas decisões administrativas inteiramente automatizadas, tendentes a melhorar a eficiência da ação pública, e a obrigação de supervisão que, reintroduzindo a intervenção humana, reduz a performance económica, mas garantem a confiança na administração pública e nas novas tecnologias, em linha com a perspetiva da União Europeia.

Com a preocupação na garantia dos direitos fundamentais a UE tem dedicado especial atenção aos elementos-chave do quadro regulatório para a IA na Europa, que deve criar um "ecossistema de confiança único", partindo de uma "abordagem centrada

no ser humano" [48]. Situando-se, assim, na linha de anteriores comunicações da Comissão relativas ao tema da IA – cuja mais recente, de 2019, intitulada "Aumentar a confiança numa inteligência artificial centrada no ser humano" [49] é disso paradigma.

[48] Livro Branco sobre a inteligência artificial - Uma abordagem europeia virada para a excelência e a confiança, Bruxelas, 19.02. 2020, COM (2020) 65 final.
[49] Aumentar a confiança numa inteligência artificial centrada no ser humano, Bruxelas, 8.4.2019, COM (2019) 168 final.

Considerações finais

O aumento crescente das possibilidades técnicas, torna a automatização de processos cada vez mais atrativa para a AP. Devido aos avanços da IA, hoje podem ser automatizados processos, que há apenas alguns anos atrás tinham de ser realizados por humanos. Mas nem todos os processos administrativos podem ser automatizados do ponto de vista técnico e/ou jurídico. De entre a multiplicidade das várias centenas de procedimentos administrativos, os decisores devem selecionar os processos considerados adequados para automatização parcial ou total.

Sim, é possível, de forma gradual, a implementação de sistemas de automatização dos atos administrativos. Mas, mais importante e urgente é colocarmos a IA ao serviço da AP, como suporte à tomada de decisão, seja nos processos mais simples e massificados, seja no apoio às decisões que carecem de tratamento de enormes volumes de informação.

Sim é possível, mas através de um processo de tomada de decisão centrado na pessoa e não na máquina!

A confiança nos processos transformacionais faz-se através da ação e pela experimentação concertadas, por isso, recomendamos: a necessidade de trabalhar um ecossistema - "end-to-end", para a implementação de um modelo de inteligência colaborativa na AP, convocando todos os intervenientes relevantes: legislativo, passando, executivo, judicial, e outros, pois só desta forma serão alcançadas mais valias, em termos de eficiência, eficácia e confiança; a criação de um novo "modelo de procedimento administrativo tecnológico", mais equipado para receber sistemas de automatização e de suporte à decisão; o desenvolvimento de novas competências e papeis na era da IA, sobretudo na AP; e, por fim, cuidar de uma regulamentação exaustiva e promotora destas novas ferramentas, tão impactantes nas nossas vidas e nos nossos direitos.

Uma coisa é certa: tudo o que puder ser automatizado será. Trata-se de tirar partido de uma tecnologia para gerar eficiência em larga escala. A utilização de soluções de IA será dominante e passará a estar presente em todos os contextos da sociedade. A questão mais importante remete aos objetivos que definirmos e à transferência dos mesmos para os sistemas de IA.

O desafio é garantir confiança, segurança jurídica e corresponsabilização!

Bibliografia

Abeliuk, A., Benjamin, D., Morstatter, F., & Galstyan, A. (2020). Quantifying machine influence over human forecasters. *Scientific Reports*, 10, 15940. https://doi.org/10.1038/s41598-020-72690-4

Adadi, A. & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable *Artificial Intelligence (XAI). 6 IEEE* Access 52138. https://doi.org/10.1109/access.2018.2870052.

Adler, P. et al., (2018). Auditing Black-Box Models for Indirect Influence. *Knowledge and Information Systems*, 54(1), 95–122. https://doi.org/10.1007/s10115-017-1116-3

Ahn, Y. and Yu-Ru Lin (2019). FairSight: Visual Analytics for Fairness in Decision Making. *IEEE Transactions on Visualization and Computer Graphics*, 1–1. https://doi.org/10.1109/TVCG.2019.2934262.

Almeida, Mário Aroso (2021). Teoria Geral do Direito Administrativo, Almedina Editora

An Ultimate Guide to Understanding Cognitive Computing, Anukrati Mehta, digitalvidya, https://www.digitalvidya.com/blog/cognitive-computing/

Ananny, M. and Kate Crawford (2016). Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. New Media and Society. 20(3). https://doi.org/10.1177/1461444816676645

Anastaspoulos, L. & Whitford, A. (2018). Machine Learning for Public Administration Research, with Application to Organizational Reputation. *arXiv*, 1805.05409. https://doi.org/10.48550/arXiv.1805.05409

Anderson, M., and Susan Leigh Anderson (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, 28(4).

Arteaga, M., Fogliato, R. & Chouldechova, A, (2020). A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).*

Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376638

Ashley 2017 - Ashley Kevin D Ashley, Artificial Intelligence and Legal Analytics (Cambridge University Press, 2017).

Ashley, K. (2017). Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316761380

Barocas, S., & Selbst, A. (2016). Big Data's Disparate Impact. *California Law Review, 104*, 671-732. http://dx.doi.org/10.15779/Z38BG31

Binns, R. (2017). Data Protection Impact Assessments: A Meta-Regulatory Approach. *International Data Privacy Law, 7*(1), 22-35.

Binns, R. (2020). Analogies and Disanalogies Between Machine-Driven and Human-Driven Legal Judgement. *Journal of Cross-disciplinary Research in Computational Law*.

Binns, R. (2020, October 07). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation and Governance, 16*, 197-211.

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? Policy Quarterly, 13(4)

Brand, D. J. (2020). Algorithmic Decision-making and the Law. *eJournal of eDemocracy and Open Government, 12*, 114-131.

Brown, Shannon (2016). Peeking inside the Black Box: A Preliminary Survey of Technology Assisted Review (TAR) and Predictive Coding Algorithms for Ediscovery. *Suffolk Journal of Trial and Appellate Advocacy*, 21(1).

Bryson, J. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*, 20. https://doi.org/10.1007/s10676-018-9448-6

Buest, R. (2017, November 03). *Artificial intelligence is about machine reasoning – or when machine learning is just a fancy plugin*. Retrieved from Digital Vertices:

https://www.cio.com/article/230943/artificial-intelligence-is-about-machine-reasoning-or-when-machine-learning-is-just-a-fancy-plugin.html

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society, 3*(1), 1-12. https://doi.org/10.1177/2053951715622512

causaLens, Truly Explainable AI: Putting the "Cause" in "Because"

Citron, D. (2008). Open Code Governance. *University of Chicago Legal Forum*, 355-387.

Cobbe, J. (2019). Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies, 39*(4), 636-655. https://doi.org/10.1017/lst.2019.9

Coglianese, C., & David, L. (2019, November 29). Transparency and Algorithmic Governance, 71. *Administrative Law Review*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3293008

Coglianese, C., & Lehr, D. (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *The Georgetown Law Journal, 105*, 1147-1223.

Cook, S., (1981). The applications of artificial intelligence to law: A survey of six current projects. *National Computer Conference*. 689–696 https://doi.org/10.1145/1500412.1500516

Courtland, R. (2018, June 21). The bias detectives: As machine learning infiltrates society, scientists grapple with how to make algorithms fair. *Nature Publishing Group, 558*, 357-360. https://doi.org/10.1038/d41586-018-05469-3

Criado, J., J. Valero & Julián Villodre, (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. *Information Polity*, 25(4), 449-470. https://doi.org/10.3233/ip-200260Corpus ID: 229326352

Daugherty, P., & Wilson, J. (2018a). Humans + Machine: Reimagining Work in the Age of AI. *Harvard Business Review Press*.

Daugherty, P., & Wilson, J. (2018b). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review Press*. 114–123

Davenport, T. H., & Kirby, J. (2016). Beyond Automation. Harvard Business Review, 58-65

Deeks, A. (2019). The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review, 119*, 1829-1850.

Desai, D., & Kroll, J. (2017). Trust but verify: A guide to Algorithms and the Law. *Harvard Journal of Law & Technology, 31*(1), 1-64.

Dong, Q., Shaogang Gong, and Xiatian Zhu (2019). Imbalanced Deep Learning by Minority Class Incremental Rectification. *Transactions on Pattern Analysis & Machine Intelligence*, 41. https://doi.org/10.48550/arXiv.1804.10851.

Doshi-Velez, F. & Been Kim (2017). Towards a Rigorous Science of Interpretable Machine Learning. https://arxiv.org/pdf/1702.08608.pdf

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Paulson, J., Gershman, S., . . . Wood, A. (2018). *Accountability of AI Under the Law: The Role of Explanation.* s.l.: Harvard Public Law. arXiv preprint arXiv:1711.01134, 2018.

Du, M., Liu, N., Hu, X. (2020). Techniques for Interpretable Machine Learning. *Communications of the ACM*, 63(1), 68-77.

Duarte de Almeida, L., (2015). Allowing for Exceptions: A Theory of Defences and Defeasibility. Law, Oxford Legal Philosophy.

Enarsson, T., Enqvist, L., & Naarttijärvi, M. (2022). Approaching the human in the loop–legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law, 31*(1), 123-153. https://doi.org/10.1080/13600834.2021.1958860

Ethics and Legal in AI: Decision Making and Moral Issues. Parliamentary Group on Artificial Intelligence. Theme Report, 27 March 2017, Big Innovation Centre.

Etscheid, J. (2019). Artificial Intelligence in Public Administration. A Possible Framework for Partial and Full Automation. In I. Lindgren, M. Janssen, H. Lee, A. Polini, M. Bolivar, H. Scholl, & E. Tambouris, *Eletronic Government* (pp. 248-261). San Benedetto: 18th International Conference on Eletronic Government. https://doi.org/10.1007/978-3-030-27325-5_19

European Group on Ethics in Science and New Technologies (2018). Artificial Intelligence, Robotics and Autonomous Systems. Brussels, 9 March 2018

Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA). https://www.darpa.mil/program/explainable-artificial-intelligence

Finck, M. (2019). Automated Decision-Making and Administrative Law. In P. Cane, H. Hofmann, E. Ip, & P. Lindseth, *The Oxford Handbook on Comparative Administrative Law* (pp. 658-676). s.l.: Oxford University Press.

Floridi, L. (2018). Soft Ethics and the Governance of the Digital

Friedler, S., Scheidegger, C., Venkatasubramanian, S.,Choudhary, S. Hamilton, Evan P. Derek Roth, D. (2018). A comparative study of fairness-enhancing interventions in machine learning. *arXiv,* 1802.04422v1 [stat.ML]. https://doi.org/10.48550/arXiv.1802.04422

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review oh Psychology, 62*, 451-482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilpin, L. H, Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., Kagal, L. (2019). Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv,* 1806.00069, https://doi.org/10.48550/arXiv.1806.00069

Grimmelmann, J. (2005). Regulation by Software. *The Yale Law Journal Company, 114*, 1719-1758.

Guidotti, R. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1-45. doi:10.1145/3236009

Harlow, C., & Rawlings, R. (2019). Proceduralism and Automation: Challenges to the Values of Administrative Law. *The Foundations and Future of Public Law*, 275-298.

Hildebrandt, M. (2011). Legal Protection by Design: Objections and Refutations. *Legisprudence, 5*, 223-248. doi:https://doi.org/10.5235/175214611797885693

Hogan-Doran, D. (2017). Computer Says "No": Automation, Algorithms and Artificial Intelligence in Government Decision-making. *The Judicial Review*, 1.

Holzinger, A. (2016). Interactive Machine Learning (iML). Informatik-Spektrum, 39, 64–68. https://doi.org/10.1007/s00287-015-0941-6

Holzinger, A., Plass, M., Holzinger, K., Crisan, G., Pintea,, C.-M., & Palade, V. (2017). A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. *IML Interactive Machine Learning*, 1-26.

Horty, J., (1997). Nonmonotonic Foundations for Deontic Logic. In Defeasible Deontic Logic, Donal Nute editors.

House of Lords Select Committee on Artificial Intelligence. (2018). *AI in the UK: ready, willing and able?*. House of Lords.

Huggins, A. (2020). Executive power in the digital age: Automation, statutory interpretation and administrative law. *In Boughey, Janina & Burton Crawford, Lisa (Eds.) Interpreting executive power. Federation Press, Australia, pp. 111-128*. Retrieved from https:// eprints.qut.edu.au/180784/

Information Commissioner's Office (2017). Big Data, Artificial Intelligence, Machine Learning and Data Protection, Version 2.2, March 2017.

Johnson, D. G. (2015). Technology with No Human Responsibility?. *Journal of Business Ethics*, 707–715. https://doi.org/10.1007/s10551-014-2180-1

Jordan, M., T. Mitchell (2015). Machine learning: Trends, perspectives, and prospects. Science, 349.

Kailash, J. (2016). *Expert Systems and Applied Artificial Intelligence.* Saint-Louis: University of Missouri. http://www.umsl.edu/~joshik/msis480/chapt11.htm.

Kang, J., & Cuff, D. (2005). Pervasive Computing: Embedding the Public Sphere. *Washington and Lee Law Review*, 62(4).

Kaplan, J., (2016). Artificial Intelligence: What Everyone Needs to Know. Oxford University Press. ISBN 9780190602390

Kellogg, K., Valentine, M., & Christin, A. (2020). Algorithms at work: the new contested terrain of control. *Academy of Management Annals, 14*(1), 366-410. https://doi.org/10.5465/annals.2018.0174

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. *Computing Research Repository (CoRR)*, 1-24.

Knight, W. (2017). The Dark Secret at the Heart of AI. *MIT Technology Review*

Kroll, J., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D., & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review, 165*, 660-661.

Lacave, C., Manuel Luque and Francisco Javier D´ıez (2007). Explanation of Bayesian Networks and Influence Diagrams in Elvira In *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems*, Man, and Cybernetics Society · September 2007. https://doi.org/10.1109/TSMCB.2007.896018. https://ieeexplore.ieee.org/document/4267869

Lamego, J. (2021). Elementos de metodologia jurídica. Coimbra: Almedina Editora.

Lee, T., Johnson, J. & Cheng, S. (2016). An Interactive Machine Learning Framework. Boston, Massachusetts, USA. *arXiv*, 1610.05463v1.

Lehr, D., & Ohm, P. (2017). Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *University of California Davis Law Review, 51*, 653-717.

Lei n.º 58/2019, de 8 de Agosto.

Lember, K. (2019). The Use of Artificial Intelligence in Administrative Acts. Juridica

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turgin Institute*, 1-97. Retrieved from https://doi.org/10.5281/zenodo.3240529

Lessig, L. (1995). The Path of Cyberlaw. *Yale Law Journal, 1743*, 1744-1745.

Liao, Q., Daniel Gruen and Sarah Miller (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 21, 2020, 1–15. https://doi.org/10.1145/3313831.3376590.

Licklider, J. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*, 4-11.

Lockwood, T. (2018). Artificial intelligence can now explain its own decision making. DataDrivenInvestor

Loi, M., Müller, A., & Spielkamp, M. (2021b). Automated Decision-Making Systems in the Public Sector. An Impact Assessment Tool for Public Authorities. *Algorithm Watch*.

MatthiasS, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information *Technology*, 6(3), 175–183. https://doi.org/10.1007/ s10676-004-3422-1

McCarthy, Daniel R. (2011). Open Networks and the Open Door: American Foreign Policy and the Narration of the Internet'. *Foreign Policy Analysis*, 7(1), 89–111.

Mehta, A., (2019). An Ultimate Guide to Understanding Cognitive Computing. *Data Science*.

Mendes, Paulo de Sousa, (2020). A representação do Conhecimento Jurídico, Inteligência Artificial e os Sistemas de Apoio à Decisão Jurídica in *Inteligência & Direito*, Manuel Lopes Rocha e Rui Soares Pereira (Coord.), Editora Almedina

Mitrou, L., Janssen, M., & Loukis, E. (2021). Human Control and Discretion in AI-driven Decision-making in Government. *14th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2021)*, (pp. 10-16). s.l.

Mittelstadt, B., (2016). Auditing for Transparency in Content Personalization Systems. *International Journal of Communication*, 10, 4991–5002

Moles, R. (1992). Expert Systems - The Need for Theory, in C.A.F.M. Grutters, J.A.P.J. Breuker, H.J. van den Herik, A.H.J. Schmidt, C.N.J. de Vey Mestdagh (eds.), *Legal Knowledge Based Systems: Information Technology & Law*, JURIX '92, Koninklijke Vermande, Lelystad, NL, 1992

Morison, J., & Adam Harkens. (2019). "Re-engineering justice? Robot judges, computerised courts and(semi) automated legal decision-making". *Legal Studies, 39*(4), 618-635. https://doi.org/10.1017/lst.2019.5

Mulholand, C., & Frajhof, I. (2019). Inteligência artificial e a lei geral de proteção de dados pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning. In A. Frazão, & C. Mulholand, *Inteligência Artificial e Direito.* São Paulo: Thomas Reuters Brasil.

Nagenborg, M., Capurro, R., Weber, J., & Pingel, C. (2008). Ethical regulations on robotics in Europe. *AI & Society*, 22 (3), 349–366. https://doi.org/10.1007/s00146-007-0153-y

O'Neil, Cathy. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

Okamoto, K. (2009). Teaching Transactional Lawyering. *Drexel Law Review*, 1(69).

Oswald, M. (2018a, September 13). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 1471-2962. https://doi.org/10.1098/rsta.2017.0359

Oswald, M., Grace, J., Urwin, S., & Barnes, G. (2018b). Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'experimental' proportionality. *Information and Communications Technology Law, 27*(2), 1-27.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information.* Cambridge: Harvard University Press.

Pasquale, F. (2016). Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2779270

Pearl, J. & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Basic Books.

Pearl, J. & Mackenzie, D. (2019). The Book of Why. The New Science of Cause and Effect. Penguin. ISBN: 9780141982410

Pilving, I., & Mikiver, M. (2020). A Kratt as an Administrative Body: Algorithmic Decisions and Principles of Administrative Law. *Juridica International*, 29, 47-61. https://doi.org/10.12697/ji.2020.29.05

Prakken, H., (1997). Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law. Law and Philosophy Library, vol. 32. Kluwer Academic Publishers, Dordrecht, Boston, and London, 1997. https://doi.org/10.1007/978-94-015-8975-8;

Qualcomm Technologies, (September 6, 2022). Is causality the missing piece of the AI puzzle? Qualcomm AI Research explores fundamental research to combine causality with AI. Qualcomm Technologies.

Redondo, M., (1997). Teorias del Derecho e Indeterninación Normativa, Doxa. *Cuadernos de Filosofía del Derecho*. 20.

Ribeiro, M., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. *22nd ACM Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining*, (pp. 1135-1144). s.l.

Robert, S., Bttner, S., Rcker, C., & Hilzinger, A. (2016). Reasoning under uncertainty: Towards collaborative interactive machine learning. In A. Holzinger, *Machine Learning for Health Informatics: State-of-the-Art and Fu-turue Challenges* (pp. 357-376). Cham: Spinger International Publishing. doi:doi: 10.1007/978-3-319-50478-0 18

Rosnay, M., (2016). Algorithmic Transparency and Platform Loyalty or Fairness in the French Digital Republic Bill. *Media Policy Project*. http://blogs.lse.ac.uk/mediapolicyproject/2016/04/22/algorithmic-transparency-and-platform-loyalty-or- fairness-in-the-french-digital-republic-bill/.

Rudin, C. (2019). Please Stop Explaining Black Box Models for High-Stakes Decisions. *Nature Machine Intelligence*, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Santoro, M., Marino, D., & Tamburrini, G. (2008). Learning robots interacting with humans: From epistemic risk to responsibility. *AI & Society*, 22(3), 301–314. https://doi.org/10.1007/ s00146-007-0155-9

Santos, H. L. (2022). Inteligência artificial e processo penal. Novas Causas, Edições Jurídicas.

Sarkar, D., (2018). The Importance of Human Interpretable Machine Learning. *Towards Data Science.* https://towardsdatascience.com/human-interpretable-machine-learning-part-1-the-need- and-importance-of-model-interpretation-2ed758f5f476 [https://perma.cc/4XD8-F7CD].

Sathya, R. & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2). https://doi.org/10.14569/IJARAI.2013.020206

Scantamburlo, T., Charlesworth, A., & Cristianini, N. (2019). Machine Decisions and Human Consequences. *ArXiv*, 1811.06747 [Cs]. http://arxiv.org/abs/1811.06747.

Scantamburlo, Teresa, Schirner, G., Erdogmus, D., Chowdhury, K., & Padir, T. (2013). The future of human-in-the-loop cyber-physical systems. *Computer, 46*(1), 36-45.

Schmelzer, R. (2020, January 9). Going Beyond Machine Learning to Machine Reasoning. *Forbes*.

Selbst, A., & Solon Barocas. (2018). "The Intuitive Appeal of Explainable Machines". *Fordham Law Review, 87*(3), 1085-1139. Retrieved from https://ir.lawnet.fordham.edu/flr/vol87/iss3/11

Sharma, S. (2022). Gartner picks emerging technologies that can drive differentiation for enterprises. Venture Beat.

Shekhar, Gaurav (2022). Causal AI — Enabling Data-Driven Decisions. *Towards Data Science*, May 26, 2022

Shyu, C.-R., Brodley, C., Kak, A., Kosaka, A., Aisen, A., & Broderick, L. (1999). ASSERT: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer Vision and Image Understanding, 75*(12), 111-132. https://doi.org/10.1006/cviu.1999.0768

Skitka, L., Kathleen, M., & Burdick, M. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies* 52, 701. https://doi.org/10.1006/ijhc.1999.0349

Sørmo, F., Cassens, J., & Aamodt, A. (2005). Explanation in Case-Based Reasoning–Perspectives and Goals. *Artificial Intelligence Review* 24, 109–143. https://doi.org/10.1007/s10462-005-4607-7

Stevens, Y., (2017). The Promises and Perils of Artificial Intelligence. Why Human Rights and the Rule of Law Matter. https://medium.com/@ystvns/the-promises-and-perils-of-artificial-intelligence-why- human-rights-norms-and-the-rule-of-law-40c57338e806, September 5, 2017.

Su, A., (2022). The Promise and Perils of International Human Rights Law for AI Governance *Law, Technology and Humans*, 4(2), 166-82. https://doi.org/10.5204/lthj.2332.

Sun, T., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36*(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Super, D. (2005). Are Rights Efficient? Challenging the Managerial Critique of Individual Rights. *California Law Review, 93*(4), 1051-1142. https://www.jstor.org/stable/3481467

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Communications of the ACM*, 56(5), 44–54. https://doi.org/10.1145/2447976.2447990

Tufekci, Z., Jillian C. York, Ben Wagner, and Frederike Kaltheuner (2015). The Ethics of Algorithms: From Radical Content to Self-Driving Cars. Berlin, Germany: European University Viadrina. Retrieved. https://cihr.eu/publication-the-ethics-of-algorithms/.

van Acken, J. (2019). Key Points of the Federal Government for a Strategy Artificial Intelligence. https://doi.org/10.13140/RG.2.2.27019.34082

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. In K. Yeung, & M. Lodge, *Algorithmic Regulation* (pp. 121-149). Oxford: Oxford University Press.

Vérine, A., & Stéphan, M. (2019). Interpretability of machine learning: what are the challenges in the era of automated decision-making processes?. Wavestone

Vincent, C. & Camp, J. (2004). Looking to the Internet for Models of Governance, 6 E*thics & Info. Tech*. (explaining that automated processes remove transparency).

Wachter (2017). Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics, 2*(6), 3-6.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDP. *Harvard Journal of Law & Technology*, 31(2). https://arxiv.org/abs/1711.00399

Watson, D. & Floridi, L. (2021). The explanation game: a formal framework for interpretable machine learning.

Williams, R. (2021). Rethinking Administrative Law for Algorithmic Decision Making. *Oxford Journal of Legal Studies, 42*(2), 468-494.

Williams, R., & Melham, T. (2020). *Automated Decision-Making in the Public Sector.* s.l.: Practical Law.

Wróblewski J., (1983). Fuzziness of legal system. Finnish Lawyers' Society.

Xiong, Momiao (2022). Artificial Intelligence and Causal Inference. Chapman & Hall

Yeung, K., (2015). Hypernudge: Big Data as a mode of regulation by design. Information, *Communication & Society*. *The Social Power of Algorithms*, 20(1). 118-136. https://doi.org/10.1080/1369118X.2016.1186713

Zalnieriute, M., Bennett Moses, L., & Williams, G. (2019). The Rule of Law and Automation of Government Decision-Making. *Modern Law Review*. 1, 8-11.

Zarsky, T. (2016). The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values*, 41(1). https://doi.org/10.1177/01622439156055