El tractament de dades en entorns digitals: els desafiaments jurídics de la intel·ligència artificial (IA)

PID_00270380

Alessandro Mantelero

Temps mínim de dedicació recomanat: 2 hores

Universitat Oberta de Catalunya

Alessandro Mantelero

Professor titular de Dret Privat i de *Data Ethics and Protection* del Politècnic de Torí. És expert científic del Consell d'Europa per a la protecció de dades i per al Consell d'Europa (*Guidelines on personal data in a world of Big Data*, 2017; *Report on Artificial Intelligence and Data Protection: Challenges and Possible Remedies*, 2019; *Guidelines on Artificial Intelligence and Data Protection*, 2019). És membre coeditor de la revista *Computer Law and Security Review* i membre del consell editorial de les revistes *European Data Protection Law Review* i *IDP: Revista d'Internet, Dret i Política*.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Mònica Vilasau Solana (2020)

Primera edició: febrer 2020

Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Índex

Introducció

La intel·ligència artificial,[1] malgrat l'atenció que està despertant recentment, no constitueix un tema nou en la recerca científica i en el debat sobre les seves possibles conseqüències socials. Si bé no és aquest el lloc per a una àmplia reflexió sobre el tema, ni per a examinar la varietat d'aplicacions concretes que se solen incloure en aquesta categoria, és necessari delimitar l'objecte de la recerca i preguntar-se per què en els últims anys s'ha obert un debat a diferents nivells sobre aquest tema.

(1)A partir d'ara IA, segons l'acrònim anglès més comú.

Lectures recomanades

W. S. McCulloch i d'altres (1943). «A Logical Calculus of the Idees Immanent». *Nervous Activity. Bulletin of Mathematical Biophysics* (vol. 5, pàg. 115-133).

A. M. Turing (1950). «Computing Machinery and Intelligence». *Mind* (vol. 49, pàg. 433-460).

La definició d'intel·ligència artificial (IA)

No hi ha una definició unitària d'intel·ligència artificial (IA). El terme *intel·ligència artificial* va ser encunyat originalment per John McCarthy, un informàtic nord-americà conegut com el pare de la IA. Amb el terme *intel·ligència artificial* se solen descriure els sistemes informàtics que són capaços d'aprendre de les seves pròpies experiències i resoldre problemes complexos en diferents situacions, habilitats que abans pensàvem que eren exclusives de l'ésser humà.

Aquesta és la definició d'IA proporcionada pel Consell d'Europa:

«AI is actually a young discipline of about sixty years, which brings together sciences, theories and techniques (including mathematical logic, statistics, probabilities, computational neurobiology and computer science) and whose goal is to achieve the imitation by a machine of the cognitive abilities of a human being. Specialists generally prefer to use the exact names of the technologies actually used (which today are essentially machine learning) and are sometimes reluctant to use the term "intelligence" because the results, although extraordinary in some areas, are still modest compared to the stated ambitions.»

Council of Europe. «What's AI?» [en línia]. <www.coe.int/en/web/artificial-intelligence/what-is-ai>

Lectura recomanada

J. McCarthy; M. L. Minsky; N. Rochester; C. E. Shannon (1955). «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence» [en línia]. <www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

Al mateix temps, per a abordar aquest debat des d'una perspectiva reguladora, és necessari preguntar-se quines formes d'aplicació de la IA són raonables d'esperar en els propers anys, amb la finalitat de definir correctament l'objecte de la regulació. En aquest sentit, cal rebutjar els escenaris apocalíptics o de ciència-ficció que prefiguren una intel·ligència artificial comparable a la humana i que plantegen preguntes sobre la subjectivitat jurídica que sembla mancar de fonament amb referència al futur proper.

Aquest renovat interès per la IA es deu, com en el passat, a una nova fase de desenvolupament de la recerca en aquest camp. Aquesta fase es refereix als perfils tecnològics necessaris i funcionals per a una aplicació concreta de la IA, més que als fonaments teòrics. En els últims deu anys, diferents factors han confluït per a crear un entorn completament nou i extremadament fèrtil respecte de la IA. Cal destacar, entre aquests factors, les transferències de dades cada vegada més ràpides i potents a través de la xarxa, una capacitat d'emmagatzematge significativament més gran i la possibilitat d'utilitzar recursos computacionals potencialment il·limitats gràcies a la computació en el núvol. A més, ha de tenir-se en compte la transformació progressiva en dades de fins i tot els esdeveniments més insignificants de la vida quotidiana de la majoria dels individus.

Lectures recomanades

V. Mayer-Schönberger; K. Cukier (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think* (pàg. 78). Londres: John Murray.

M. Lycett (2013). «Datafication: making sense of (big) data in a complex world». *European Journal of Information Systems* (vol. 22, núm. 4, pàg. 381-386).

L'aprenentatge automàtic

La intel·ligència artificial inclou formes diferents d'aprenentatge automàtic. L'aprenentatge automàtic pot descriure's com un conjunt de tècniques i eines que permeten que les computadores puguin aprendre nocions o entendre fets, creant algorismes matemàtics basats en dades acumulades a gran escala que sembla que poden raonar independentment de la intervenció humana i també construir nous algorismes.

L'aprenentatge profund és una manera d'aprenentatge automàtic. Alguns tipus d'aprenentatge profund es basen en l'anomenada «xarxa neuronal», que utilitza un conjunt conegut de dades d'entrenament que ajuden els algorismes d'autoaprenentatge a realitzar una tasca. Això està condicionat al fet que la mateixa xarxa pugui determinar la resposta correcta per a resoldre la tasca. Un exemple d'aplicació d'aquestes xarxes neurals va ser el programa AlphaGo, que va derrotar un dels més grans campions del món de Go.

En el passat, les teoritzacions de la IA havien xocat amb barreres tecnològiques que limitaven les seves possibles aplicacions. Ara, els desenvolupaments recents esmentats aquí han alliberat aquest potencial. Això ha portat a un canvi, que ha permès noves formes de gestió de dades destinades a extreure nous coneixements, fins i tot de naturalesa predictiva. El *big data* i l'aprenentatge automàtic són els productes més recents d'aquest procés de desenvolupament.

Lectura recomanada

The Norwegian Data Protection Authority (2018). *Artificial Intelligence and Privacy Report* [en línia]. <www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>

En aquest sentit, hem de tenir en compte les aplicacions concretes d'aquestes tecnologies, per a entendre quin tipus d'IA poblarà les aplicacions dels propers anys, confirmant que encara estem molt lluny de l'anomenada «IA general», és a dir, d'un model d'IA capaç d'enfrontar-se a qualsevol tipus de problema i context, de manera similar al que li passa a la ment humana. En aquest moment i en els propers anys, de fet, solament ens trobarem davant formes d'IA capaces de realitzar una o més tasques específiques, certament amb una gamma variada d'aplicacions (des del reconeixement d'imatges, passant per la traducció de textos, fins a les aplicacions de jocs), però sempre amb un enfocament específic.

En enfrontar-se al tema de la regulació de la IA, és necessari enfocar el debat sobre les aplicacions concretes i que s'estan desenvolupant, sense considerar els escenaris més extrems sobre la interacció home-màquina. En aquest sentit,

els algorismes d'IA actuals ja tenen un impacte en l'ús de dades personals i plantegen interrogants sobre la idoneïtat de les regulacions existents per a abordar els problemes plantejats per aquests nous paradigmes.

Lectures recomanades

CNIL (2017). *How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence. Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill* [en línia]. <www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>

N. Bostrom (2016). *Superintelligence paths, dangers, strategies*. Oxford: Oxford University Press.

R. Kurzweil (2016). *The singularity is near: when humans transcend biology*. Londres: Duckworth.

A més, l'impacte de la IA no es refereix solament al processament de dades en si mateix, sinó a les finalitats aplicatives i als principis i valors que han de guiar la implementació de la IA en casos concrets. De fet, està sorgint una tendència cap a una societat tecnocràtica impulsada pel mercat que condueix a la monetització de les dades personals, a formes de control social i solucions «econòmiques i ràpides» de presa de decisions tant a gran escala (per exemple, en el context de les ciutats intel·ligents) com a petita escala (per exemple, medicina de precisió, assistents personals, dispositius domèstics intel·ligents, etc.). Aquesta tendència planteja desafiaments significatius per a la protecció de l'autodeterminació individual i presenta problemes crítics per als models tradicionals centrats únicament en la protecció de la informació personal.

Lectures recomanades

S. Speikermann (2016). *Ethical IT Innovation. A Value-Based System Design Approach* (pàg. 152). Boca Raton: CRC Press.

A. Mantelero (2018). *Ciudadanía y gobernanza digital entre política, ética y derecho*. A: T. Quadra Salcedo; J. L. Piñar Mañas. *Sociedad Digital y Derecho*. Madrid: Butlletí Oficial de l'Estat.

La bulímia de les dades, la complexitat del processament i una lògica profundament centrada en el mesurament dels fenòmens i les interaccions socials poden soscavar l'ús democràtic de la informació i imposar una espècie de dictadura de dades en la qual els models algorítmics es desenvolupen, s'apliquen i s'utilitzen per a la presa de decisions sense sentit crític. Per a evitar que les conseqüències adverses de la IA prevalguin sobre els beneficis, és necessari preservar i reafirmar la centralitat de l'ésser humà respecte al desenvolupament tecnològic en general i, específicament, en relació amb la IA.

Lectura recomanada

J. L. Piñar Mañas (2018). *Derecho e innovación tecnológica. Retos de presente y futuro*. Madrid: CEU Ediciones.

Això significa reafirmar el predomini dels drets fonamentals en aquest àmbit. En aquest sentit, el dret a la protecció de les dades pot convertir-se en el punt de partida per al desenvolupament d'una IA que no estigui impulsada pel mer interès econòmic o per l'eficiència dels processos, sinó que sigui capaç de combinar innovació i protecció dels drets individuals i interessos col·lectius.

Lectura recomanada

A. Rouvroy (2016). *Of Data and Men»: Fundamental Rights and Liberties in a World of Big Data* [en línia]. <rm.coe.int/CoERMPublicCommonSearch Services/DisplayDCTMContent?documentId =090000168o6a6020>

Des d'aquesta perspectiva, el desenvolupament de la IA –que se centra necessàriament en les dades personals quan es refereix a aspectes individuals i socials– ha de basar-se en una lectura crítica i actualitzada dels principis de proporcionalitat, responsabilitat i transparència, així com en formes adequades de gestió del risc i en formes de participació activa de les parts interessades.

1. Els desafiaments de la IA

Com ha succeït en altres ocasions, el canvi de paradigma introduït primer pel *big data* i després per la IA s'ha trobat amb un marc regulatori europeu sobre processament de dades que és parcialment inadequat per a proporcionar respostes oportunes.

Lectura recomanada

A. Mantelero (2016). «Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection». *Computer Law & Sec. Review* (vol. 32, pàg. 238-255).

La creixent concentració del poder d'informació, la foscor freqüent del seu ús i l'impacte de l'ús de les dades en els processos de presa de decisions a gran escala semblen ser abordats només parcialment pel legislador europeu, que, en el Reglament (UE) 2016/679, segueix sent principalment fidel a un model de protecció centrat en els drets i en el paper de la persona interessada, reafirmant principis com els de transparència i minimització, que són difícils de conciliar amb la naturalesa de la IA.

Si, d'una banda, el nou reglament mostra aquests límits, per una altra, marca un punt d'inflexió en el sentit de l'anàlisi i la gestió del risc relacionat amb el tractament. Es desplaça l'enfocament de la normativa des de l'autodeterminació de la persona interessada cap a una major centralitat de la responsabilitat del titular del tractament i d'aquells involucrats en el tractament de les dades.

En aquest sentit, la responsabilitat i el principi d'*accountability*, l'objectiu dels quals és demostrar el compliment dels principis i les obligacions establerts pel Reglament, constitueixen el nucli del nou marc europeu de protecció de dades i són un element útil per a abordar els possibles efectes negatius de l'ús de la IA.

No obstant això, la transició actual cap aquesta diferent manera de tractar la protecció de dades, centrada en el risc, sembla encara incompleta. Segueixen existint elements del model anterior centrats en la dimensió individual de la persona interessada i no hi ha hagut cap replantejament de l'arquitectura reguladora, els principis fundadors de la qual es remunten a la dècada dels noranta i, segons com, a la dels setanta.

En aquest context, les disposicions de l'article 22 del Reglament i el seu paper potencial en el context de la IA són particularment interessants en relació amb el debat doctrinal.[2] En aquest punt, però, cal assenyalar que la importància d'aquesta norma pot sobreestimar-se, ja que el nombre de casos en els quals s'adopten decisions de manera exclusivament automatitzada que produeixen «efectes jurídics» per a la persona afectada o que poden afectar «de manera similar i significativa la seva persona» segueix sent limitat. En molts casos, de fet, els sistemes d'IA són eines d'anàlisis per a fonamentar les decisions, que seran preses per les persones amb poder de decisió.

[2] Vegeu l'art. 35, Regl. (UE) 2016/679.

Lectures recomanades

M. Veale; L. Edwards (2018). «Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling». *Computer Law & Security Rev.* (vol. 34, núm. 2, pàg. 398-404).

L. Edwards; M. Vale (2017). «Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for». *Duke Law & Technology Review* [en línia]. <papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855>

L. A. Bygrave (2001). «Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling». *Computer Law & Security Rev.* (vol. 17, núm. 1, pàg. 17).

W. Schreurs; M. Hildebrandt; E. Kindt-M. Vanfleteren (2010). «Cogitas, Ergo Sum. The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector». A: M. Hildebrandt; S. Gutwirth (ed.). *Profiling the European Citizen. Cross-Disciplinary Perspective* (pàg. 241). Dordrecht: Springer.

Per tant, el problema més rellevant es refereix al grau efectiu de llibertat que caracteritza l'autodeterminació dels responsables de la presa de decisions pel que fa als suggeriments proporcionats pels algorismes. En aquest context, hem de preguntar-nos si la presència d'un subjecte que adopta decisions acaba sent més la raó formal de l'exclusió de l'aplicació de l'art. 22 que una garantia efectiva d'autonomia de judici pel que fa a la màquina.

Lectura recomanada

Article 29 Data Protection Working Party (2017). *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (pàg. 10).

També cal assenyalar que el dret a obtenir la intervenció humana no sembla ser particularment decisiu pel que fa als problemes relacionats amb l'ús de la IA. D'una banda, de fet, aquesta protecció no operarà quan es persegueixen finalitats contractuals o el tractament de dades es basa en el consentiment de la persona interessada,[3] fet que redueix la protecció proporcionada per la norma. D'altra banda, la feblesa dels responsables de la presa de decisions enfront dels suggeriments oferts pels algorismes es repeteix en aquest cas i planteja dubtes sobre l'eficàcia concreta de la «intervenció humana».

[3]Vegeu l'art. 22, par. 2, a) i c), Regl. (UE) 2016/679.

En termes d'anàlisi de risc, el RGPD conté normes importants que també podrien aplicar-se per a avaluar l'impacte de l'ús de les dades en el context de les aplicacions d'IA. No obstant això, el nou reglament no conté indicacions específiques per a definir un model d'anàlisi de riscos que sigui capaç de tenir en compte l'impacte més ampli que té l'ús de les dades en el context de la IA; un impacte que va més enllà de la simple protecció de les dades. Això es veu confirmat pels models de DPIA, que encara se centren principalment en la seguretat de les informacions i la qualitat de les dades, més que en les conseqüències sobre una pluralitat de drets i la dimensió col·lectiva de l'ús de les dades.

Lectura recomanada

A. Mantelero (2016). *Op. cit.*

A més, per la seva pròpia naturalesa, la legislació sobre les dades personals no pot proporcionar respostes completes a les preguntes sobre l'impacte social de la IA i les qüestions ètiques que sorgeixen de l'ús de la IA. Per aquesta raó, s'han proposat solucions més innovadores, com les directrius sobre *big data* adoptades pel Consell d'Europa el 2017 i les directrius sobre la IA i el tracta-

ment de dades que el Consell d'Europa va adoptar al gener del 2019, totes dues destinades a identificar noves maneres de regular l'impacte dels algorismes en la societat.

Lectures recomanades

Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data [en línia]. <rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a>

A. Mantelero (2017). «Regulating Big Data. The guidelines of the Council of Europe in the context of the European data protection framework». *Computer Law & Sec. Review* (vol. 33, pàg. 584-602).

Guidelines on Artificial Intelligence and Data Protection [en línia]. <www.coe.int/en/web/data-protection/-/new-guidelines-on-artificial-intelligence-and-personal-data-protection>

Independent High-Level Expert Group on Artificial Intelligence (2019). *Ethics Guidelines for Trustworthy AI* [en línia]. <ec.europa.eu/futurium/en/ai-alliance-consultation>

Council of Europe-Committee of Experts on Internet Intermediaries (MSI-NET) (2018). *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications* [en línia]. <rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>

European Data Protection Supervisor-Ethics Advisory Group (2018). *Towards a digital ethics* [en línia]. <edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf>

Parlament Europeu (2017). *European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI))* [en línia]. <www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//EN&language=EN>

European Union Agency for Fundamental Rights (FRA) (2018). *Big Data: Discrimination in Data-Supported Decision Making* [en línia]. <fra.europa.eu/en/publication/2018/big-data-discrimination>

Les tecnologies destinades a un ús massiu i intensiu de les dades, com la IA, representen un desafiament per a l'aplicació de diversos principis tradicionals pel que fa a la protecció de dades, i els fa més borrosos, menys clars o més difícils d'aplicar. En aquest sentit, diversos autors han assenyalat la feblesa, per exemple, de l'eficàcia del consentiment informat, com una eina per a una veritable autodeterminació informativa. Això és encara més evident enfront de les formes de creació de perfils basades en algorismes d'IA, que són moltes vegades foscos i complexos, o en pràctiques ocultes de *nudging* que soscaven la noció de control de la informació per part de les persones interessades.

Lectures recomanades

Ex multis, **M. Hildebrandt** (2016). *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing.

S. Barocas; H. Nissenbaum (2015). «Big Data's End Run around Anonymity and Consent». A: J. Lane i d'altres (dir.). *Privacy, big data, and the public good: frameworks for engagement*. Cambridge: Cambridge University Press.

D. K. Citron; F. Pasquale (2014). «The Scored Society: Due Process For Automated Predictions». *Wash. L. Rev.* (vol. 89, pàg. 1-33).

A. Mantelero (2014). «The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics». *Computer Law & Sec. Rev.* (vol. 30, pàg. 643-660).

I. S. Rubinstein (2013). «Big Data: The End of Privacy or a New Beginning?». *International Data Privacy Law* (vol. 3, núm. 2, pàg. 74-87).

També ha d'assenyalar-se que la noció d'autodeterminació, en el context de la IA, no pot considerar-se limitada al mer ús de dades, sinó que assumeix una major importància en relació amb una llibertat d'elecció més general, tant pel que fa a solucions centrades en l'ús de la IA, com a la pretensió de poder utilitzar una versió de dispositius i serveis equipats amb IA que pugui ser no «intel·ligent». Aquest tipus d'opció que exclou la IA no es refereix solament a la dimensió individual i a l'ús de dispositius/serveis específics per part de l'usuari, sinó també a la llibertat més àmplia d'una comunitat per a decidir sobre el paper que ha d'exercir la IA en la configuració de la dinàmica social, el comportament col·lectiu i les decisions que afecten grups d'individus.

Lectures recomanades

Office of the Privacy Commissioner of Canada (2016). *The Internet of Things: An Introduction to Privacy Issues with a Focus on the Retail and Home Environments* [en línia]. <www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/iot_201602/#heading-0-0-2-15>

Asilomar AI Principles (2017). <futureoflife.org/ai-principles>

La doctrina ha intentat proporcionar una resposta a aquestes demandes secundant la necessitat d'enfortir el paper de la transparència en el tractament de dades o, quant al consentiment, suggerint formes més flexibles, com el consentiment ampli i el consentiment dinàmic. Encara que cap d'aquestes solucions pot, per si sola, proporcionar una resposta exhaustiva a la crisi del consentiment en el context de la IA, en algunes àrees, aquestes solucions –soles o combinades– poden enfortir l'autodeterminació individual efectiva.

Lectures recomanades

Ex multis, **L. Edwards i d'altres** (2017). «Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For». *Duke Law and Technology Review* (vol. 16, núm. 1, pàg. 18-84).

A. Selbst i d'altres (2017). «Meaningful Information and the Right to Explanation». *International Data Privacy Law* (vol. 7, núm. 4, pàg. 233-242).

S. Wachter i d'altres (2017). «Why a right to explanation of automated decision - making does not exist in the General Data Protection Regulation». *International Data Privacy Law* (vol. 7, núm. 2, pàg. 76-99).

M. Sheehan (2011). «Can Broad Consent be Informed Consent?». *Public Health Ethics* (vol. 3, pàg. 226-235).

J. Kaye i d'altres (2015). «Dynamic consent: a patient interface for twenty-first century research networks». *European Journal of Human Genetics* (vol. 23, núm. 2, pàg. 141-146).

Un altre problema important de les aplicacions de la IA està relacionat amb els biaixos potencials que poden afectar aquests sistemes. Si d'una banda els sistemes d'IA poden contribuir a reduir o eliminar els biaixos que poden afectar les persones en el procés de presa de decisions, d'altra banda, és possible que aquests mateixos estiguin viciats per prejudicis o elements desviats que porten a conclusions errònies en el procés automatitzat de presa de decisions. Tant els models deterministes d'IA com els d'aprenentatge automàtic utilitzen dades preexistents com a base per a extreure més informacions (anàlisis de *big data*) o crear i «entrenar» models d'aprenentatge automàtic. D'aquí la centralitat del tema del biaix en la selecció i l'ús de la informació.

El biaix potencial pot afectar aspectes diferents: la metodologia de recerca, l'objecte de la recerca (aquest és el cas, per exemple, del biaix social a causa del biaix en les sèries històriques de dades o a la falta de representació adequada d'algunes categories), les fonts de dades (biaix en els processos de selecció de dades) o el mateix comportament de l'autor de l'activitat de recerca.

Lectura recomanada

M. Veale; R. Binns (2017). «Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data>» [en línia]. *Big & Data Society* (vol. 4, núm. 2). <doi.org/10.1177/2053951717743530>

La presència de biaix pot afectar negativament el desenvolupament i l'aplicació d'algorismes, amb un impacte encara més gran en el cas d'algorismes d'aprenentatge automàtic, on els biaixos poden influir tant en el disseny d'aquests com en el desenvolupament. D'aquí la necessitat d'abordar un enfocament des del disseny per a evitar «potential hidden data biases and the risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects, in both the collection and analysis stages». No obstant això, aquest enfocament no es limita a la simple privadesa o protecció de dades des del disseny, sinó que ha de tenir en compte cada vegada més l'àmplia varietat de drets, llibertats i interessos potencialment afectats per l'ús de solucions centrades en la intel·ligència artificial.

Lectura recomanada

A. Mantelero (2019). *Report on Artificial Intelligence and Data Protection: Challenges and Possible Remedies* (pàg. 9).

Per tant, davant la complexitat dels sistemes d'IA i la pluralitat de possibles impactes, tant respecte a les persones com respecte a la societat, es fa evident el risc de deixar la tasca de la realització d'aquests sistemes als desenvolupadors de programari. En conseqüència, no és possible deixar l'avaluació i gestió d'aquests impactes a les decisions dels desenvolupadors ni a la visió del món que ells o les companyies a les quals pertanyen tenen.

Lectura recomanada

AI Now Institute (2017). *AI Now 2017 Report* (pàg. 18) [en línia]. <assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf>

D'aquí la necessitat d'un desenvolupament més unànime i participat de la IA, que vegi un paper actiu de les diferents categories de parts interessades i tingui en compte coneixements diferents.[4] D'aquesta manera es poden tancar les bretxes cognitives i experiencials d'una formació tècnica, que és adequada per a desenvolupar el programari, però no proporciona mesures suficients i adequades per a interpretar les conseqüències socials de l'ús dels algorismes d'IA.

[(4)]Vegeu l'art. 35, par. 9, Regl. (UE) 2016/679.

Lectura recomanada

Article 29 Data Protection Working Party (2018). *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is «likely to result in a high risk» for the purposes of Regulation 2016/679*.

Com es va esmentar, una altra manera de reduir el possible biaix d'aplicació de la IA és recórrer a formes col·laboratives d'avaluació de riscos centrades no solament en la seguretat i la qualitat de les dades, sinó també en la participació activa de grups potencialment afectats per les aplicacions d'IA. Així mateix, han de participar-hi les parts interessades capaces de contribuir a la identificació i eliminació dels biaixos i els potencials impactes negatius de l'ús de dades. Aquest enfocament, dirigit al desenvolupament responsable de solucions basades en IA, té com a objectiu evitar factors de distorsió que poden afectar els conjunts de dades o els algorismes.

En un context necessàriament caracteritzat per un nivell significatiu de complexitat i una transparència limitada dels sistemes, el recurs a formes d'anàlisis preliminars de riscos i planificació responsable sembla, de fet, prometre una eficàcia més gran de la que es pot esperar de qualsevol remei *a posteriori*, adoptat una vegada que s'hagin produït efectes discriminatoris. De la mateixa manera, la resposta centrada en la gestió del risc sembla més eficaç que aquella basada en la transparència dels algorismes. I això perquè aquesta transparència sembla difícil d'aconseguir i, sovint, no resulta molt efectiva perquè existeix una falta d'interès i de coneixement per part dels afectats.

Lectures recomanades

A. D. Selbst (2017). «Disparate Impact in Big Data Policing». *Georgia Law Review* (vol. 52, núm. 1, pàg. 109-195).

R. Brauneis; E. P. Goodman (2018). «Algorithmic Transparency for the Smart City». *Yale J. L. & Tech.* (vol. 20, pàg. 131).

Aquest enfocament *ex ante* hauria d'induir, per tant, una reflexió més profunda sobre els conjunts de dades utilitzades per a la creació i l'entrenament dels algorismes amb la finalitat d'evitar conseqüències desfavorables derivades, per exemple, del denominat biaix històric resultant de l'ús de conjunts de dades preexistents. Aquest risc s'agreuja quan els conjunts de dades no es creen *ad hoc* per al desenvolupament de l'aplicació específica, sinó que consisteixen en bases de dades creades per a diferents propòsits o disponibles al mercat, la qual cosa no permet als desenvolupadors un coneixement precís dels criteris de composició de la base de dades.

Com es poden reduir els riscos

Quant a això, amb la finalitat de reduir aquests riscos, s'ha suggerit introduir formes de traçabilitat de les fonts, desenvolupament i ús dels conjunts de dades, que s'utilitzen per a aquest fi, al llarg del seu cicle de vida.

Biaix històric

Això és el biaix que es genera quan una sèrie històrica determinada presenta una tendència adversa respecte a algunes categories, per la qual cosa el desenvolupament de l'algorisme basat en l'ús de dades històriques té una alta probabilitat d'adquirir i mantenir aquesta tendència com a element caracteritzador.

Sempre tenint en compte les solucions enfocades en el disseny de desenvolupament d'IA, és possible introduir proves precises en la fase d'entrenament dels algorismes abans de la seva aplicació *in vivo*. No obstant això, en alguns casos, l'ús d'una menor quantitat de dades necessari en la fase d'entrenament no permet als algorismes d'aprenentatge automàtic predir els efectes distorsionadors que es produeixen quan aquests s'utilitzen a gran escala. En aquest sentit, a diferència del que passa en el context de l'ús de dades personals amb finalitats estadístiques, cal assenyalar que els biaixos de les aplicacions d'IA poden originar-se en les mateixes solucions tecnicoinformàtiques adoptades, i no solament en el comportament o els errors dels científics de dades.

Lectures recomanades

Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data, op. cit.

M. L. Cummings i d'altres (2018). *Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated*. Londres: Chatham House. The Royal Institute of International Affairs.

R. Caruana i d'altres (2015). «Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission». A: *Proceedings of the 21st Annual SIGKDD International Conference on Knowledge Discovery and Data Mining* (pàg. 1721-1730).

A més, en el context de la IA, l'avaluació del biaix potencial també pot ser controvertida. La multiplicitat de les variables involucrades i la classificació de les persones en grups que no corresponen necessàriament a les categories tradicionals que s'utilitzen en els casos de discriminació provoquen que sigui més difícil identificar els prejudicis potencials.

Lectures recomanades

J. Donovan i d'altres (2018). *Algorithmic Accountability: A Primer* [en línia]. <datasociety.net/output/algorithmic-accountability-a-primer>

A. Mantelero (2016). *Op. cit.*

Finalment, hem de considerar la línia argumentativa dirigida a disminuir el possible biaix, i els conseqüents efectes perjudicials, de la IA sobre la base de la fal·libilitat que ja afecta la presa de decisions per part d'éssers humans. En aquest sentit, s'argumenta que la fal·libilitat humana pot ser reduïda recorrent a la IA, ja que elimina comportaments negligents o irracionals. No obstant això, quatre arguments diferents s'oposen a aquesta conclusió.

En primer lloc, les solucions d'IA estan destinades a una aplicació en sèrie, de la qual cosa es dedueix que, com en el cas de la responsabilitat per productes defectuosos, les solucions qualitativament deficients afecten inevitablement una pluralitat de persones que comparteixen connotacions iguals o similars, mentre que l'error humà en el procés de presa de decisions afecta necessàriament només el cas específic que està decidit.

En segon lloc, encara que hi ha àrees en les quals els índexs d'error de la IA són propers o inferiors als índexs d'error humà (per exemple, el reconeixement d'imatges), en molts altres contextos hi ha marges d'error excessius i inacceptables que fan que sigui indesitjable la transició des de la fal·libilitat humana a la fal·libilitat algorítmica.

En tercer lloc, no ha de passar-se per alt la dimensió sociocultural de l'error humà en termes d'acceptabilitat social. En l'àmbit polític i social, és molt més difícil admetre la fal·libilitat dels algorismes en els quals es basen precisament les solucions per a evitar decisions afectades per biaixos o altres formes de perjudicis que la humana.

Lectura recomanada

Council of Europe. *Guidelines on Artificial Intelligence and Data Protection, op. cit.*

Finalment, fins i tot admetent la possibilitat de comparar les decisions humanes amb les que pot prendre un algorisme, aquesta mateixa comparació resulta ser metodològicament difícil. En particular, no es pot fer una simple comparació quantitativa basada en el nombre d'errors i les seves conseqüències (per exemple, el nombre mitjà de víctimes causades per automòbils amb conductor humà en comparació del mateix nombre en el cas de vehicles autònoms). De fet, en avaluar les conseqüències de la IA i les decisions humanes, és necessari tenir en compte la distribució dels efectes (és a dir, les persones afectades negativament que pertanyen a diferents categories, les diferents condicions en les quals es va produir el dany, la gravetat de les conseqüències, etc.). A més, aquest tipus d'enfocament quantitatiu, que considera positivament la solució que ofereix el menor impacte en termes d'error, sembla estar en contrast amb l'enfocament precautori, que requereix l'adopció de polítiques actives de prevenció de riscos, més que una mera reducció de danys.

2. Els límits de la transparència

El recurs a obligacions de transparència sovint s'invoca com una eina per a resoldre, almenys en part, qüestions crítiques esmentades en els paràgrafs anteriors. Des d'aquest punt de vista, la transparència faria que els subjectes interessats fossin més conscients, per tal de mitigar els límits que afecten l'autodeterminació individual pel que fa al tractament de dades en el context de la IA. La transparència ajudaria llavors a aclarir els propòsits d'aquests tractaments i, finalment, però no menys important, serviria per a prevenir qualsevol biaix que pogués ser una font de prejudicis potencials, especialment en termes de discriminació.

Abans de reflexionar sobre com, en la pràctica, aquestes expectatives poden materialitzar-se solament en part pel que fa a l'adopció de polítiques de transparència, cal destacar que la mateixa noció de transparència en el context dels algorismes té contorns sovint indefinits, la qual cosa dificulta fins i tot invocar una eficàcia operacional significativa.

La noció de transparència pot, de fet, tenir diferents significats. Pot interpretar-se de manera que les persones interessades siguin conscients del fet que les solucions d'IA s'utilitzen en la interacció amb elles mateixes, o pot entendre's com a transparència del processament de les dades que les concerneixen. També pot implicar la descripció de la lògica dels algorismes utilitzats per a aquest propòsit o, finalment, arribar a un accés directe a l'estructura de l'algorisme i, quan correspongui (casos d'aprenentatge automàtic), als conjunts de dades utilitzades per a la fase d'entrenament de l'algorisme.

Lectura recomanada

R. Binns i uns altres (2018). «It's Reducing a Human Being to a Percentage, in Perceptions of Justice» [en línia]. *Algorithmic Decisions*, ArXiv:1801.10408 [Cs] (pàg. 1-14). <doi.org/10.1145/3173574.3173951>

Examinant aquí breument els diferents aspectes de la noció de transparència en el context de la IA, cal assenyalar que, si bé és important tenir un control públic sobre els models automatitzats de presa de decisions, la primera perspectiva, que consisteix en el simple coneixement sobre l'ús de la IA, és útil perquè les persones interessades prenguin consciència del context, però és poc eficaç a l'hora d'abordar els riscos d'un possible ús il·legítim de dades i de la IA.

Lectura recomanada

D. Reisman i d'altres (2018). *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* [en línia]. <ainowinstitute.org/aiareport2018.pdf>

En l'altre extrem, l'accés a l'estructura de l'algorisme i, quan correspongui, a les dades utilitzades per a l'entrenament sens dubte serà efectiu per a detectar possibles biaixos o l'ús il·legítim de la informació. No obstant això, una noció tan àmplia de transparència amb freqüència entraria en conflicte amb els drets de propietat intel·lectual i també podria implicar qüestions relacionades amb la competència i la llibertat d'empresa.

A més, si aquests obstacles no existissin o poguessin superar-se (per exemple, a través d'un accés restringit a autoritats de control o judicials, com sol ser el cas), en qualsevol cas es donaria el fet que la complexitat dels models adoptats

constitueix sovint un desafiament considerable per a les habilitats cognitives humanes, la qual cosa implica que els processos implementats per la màquina siguin inexplicables.

Finalment, cal assenyalar que hi ha casos en els quals l'ús de la transparència pot suposar altres dificultats. Es pot revelar com un obstacle per a aconseguir els objectius propis de l'Administració pública en l'exercici de les seves competències. Per exemple, en les tasques relacionades amb la prevenció del delicte (pensem en com la transparència dels algorismes dels sistemes de policia predictius podria soscavar-ne l'eficàcia). En altres casos, l'accés a l'algorisme entra en conflicte amb les obligacions de seguretat del responsable del tractament respecte a les dades personals de les persones interessades diferents d'aquelles que sol·liciten l'accés, fet que exposa aquesta informació a possibles riscos en termes de seguretat informàtica.

Per aquestes raons, la solució intermèdia entre les dues analitzades anteriorment, que consisteix en el coneixement de la lògica de l'algorisme utilitzat, sembla la més practicable. Així i tot, la transparència dels algorismes es pot interpretar més o menys estrictament. Proporcionar informació sobre les dades d'entrada i els resultats esperats, descriure les variables utilitzades i el pes que se'ls atribueix, o explicar el model analític utilitzat, són les diverses maneres que pot assumir la noció de transparència pel que fa a la lògica dels algorismes d'IA.

L'aprenentatge supervisat i sense supervisió

L'aprenentatge automàtic generalment comença amb informació seleccionada que conté *patterns* o similituds, després el programari d'aprenentatge automàtic identifica els *patterns* que es troben en aquesta informació i, finalment, genera un model que pot reconèixer els *patterns* que sorgeixen en dades noves.

Hi ha diverses maneres d'aprenentatge, que poden utilitzar dades amb etiqueta o no. Les dades etiquetades són dades que inclouen informacions sobre el contingut de cada dada (en el cas d'imatges, les etiquetes poden ser, per exemple, la raça, el gènere o el sexe d'un gos que és a la imatge). Aquestes dades etiquetades s'utilitzen en l'aprenentatge supervisat, on les etiquetes fan possible la supervisió mitjançant la qual l'algorisme s'entrena amb un «històric» de dades i aprèn a assignar les etiquetes a les noves dades que se li faciliten. En aquests casos s'entrena una màquina d'IA amb un resultat esperat i si la sortida generada per la màquina és incorrecta, el procés es repeteix sobre el mateix conjunt de dades fins que aquesta no tingui més errors. És evident que les característiques de les dades etiquetades són decisives per a una categorització correcta i per a produir resultats correctes.

En el cas de l'aprenentatge sense supervisió s'utilitzen dades que no s'han etiquetat prèviament, i el sistema ha d'agrupar les dades que siguin similars. En aquest cas s'utilitza un conjunt de dades en el qual l'algorisme ha de buscar agrupaments basats en similituds, però res no garanteix que aquestes tinguin algun significat o utilitat.

Independentment dels algorismes o mètodes utilitzats per a l'aprenentatge automàtic, el resultat serà un model que pot ser alimentat amb noves dades per a produir el tipus de resultat desitjat. Això pot ser, per exemple, una classificació o un grau de probabilitat.

No obstant això, fins i tot en aquest cas, els models d'anàlisis complexes que se centren en la IA, com els algorismes d'aprenentatge profund, constitueixen un desafiament per a aquesta noció de transparència, entesa com una explicació

de la lògica dels algorismes, atès que els sistemes no determinístics dificulten el subministrament d'informació detallada sobre la lògica existent darrere del tractament de les dades.

A més, la naturalesa dinàmica de molts algorismes contrasta amb la mateixa noció de transparència, que per la seva pròpia naturalesa és estàtica, una «fotografia» d'una situació en un moment donat. De fet, l'actualització i modificació contínua dels algorismes d'IA provoquen que sigui difícil fer-ne l'avaluació sobre la base d'una anàlisi que requereix temps per a la seva realització i els resultats de la qual, per tant, es poden aconseguir quan l'algorisme ja ha assumit una estructura diferent.

Aquesta última observació també destaca com la mera transparència no és suficient per a aclarir l'impacte dels algorismes i com, en canvi, l'accés als algorismes d'IA ha de ser seguit per una anàlisi precisa, i sovint complexa, per a verificar els riscos de possibles prejudicis. Per tant, és necessari disposar de recursos adequats en termes de tecnologies, temps i competències, que no permeten obtenir els beneficis de la transparència de manera immediata i en totes les situacions i per part de totes les persones interessades.

Certament, es pot al·legar que el periodisme de recerca, les associacions que protegeixen els interessos de categories específiques (per exemple, minories o treballadors), així com el paper de les autoritats de control, podrien tenir un paper més rellevant i reduir aquests problemes. Si bé això és cert, també ho és que l'ús de solucions d'IA és cada vegada menys costós i, en conseqüència, més generalitzat i capil·lar. D'aquí la dificultat de poder imaginar que aquestes diferents entitats i formes de vigilància poden supervisar i controlar de manera eficaç totes les aplicacions d'IA.

Aquests factors limiten les solucions operatives centrades en els mecanismes d'auditoria o en la intervenció humana en el procés de la presa de decisions. En resposta a aquests problemes, la recerca més recent té com a objectiu desenvolupar algorismes capaços de supervisar altres algorismes, que permeten la detecció automàtica dels biaixos potencials. En aquest sentit, encara que la intenció de desenvolupar tecnologies responsables és apreciable, existeixen dubtes quant al fet que la introducció d'una espècie de supervisor algorítmic es tradueixi en una reducció de la complexitat de la governança de les dades, especialment assumint la perspectiva de les parts interessades.

Lectura recomanada

M. Veale; R. Binns (2017). *Op. cit.*

No obstant això, els assumptes crítics remarcats aquí pel que fa al paper de la transparència no han d'entendre's com un intent d'afeblir els arguments a favor d'una major transparència o de disminuir aquest principi des d'una perspectiva de salvaguarda de l'autodeterminació de les persones interessades, sinó més aviat com una exhortació a comprendre plenament la complexitat d'aquest principi, fins i tot en un pla operatiu.

En tot cas, ha de tenir-se en compte com la transparència és solament una part de la solució als desafiaments de la IA i té diverses limitacions que han d'abordar-se completament. Així mateix, l'algorisme, la transparència del qual es discuteix habitualment, és solament un dels components d'una aplicació d'IA, mentre que d'altres, igualment rellevants, són els conjunts de dades utilitzades per a crear i entrenar l'algorisme i per a les dades utilitzades en la posterior fase de processament i anàlisi. Per tant, independentment de la transparència de l'algorisme, aquests conjunts de dades poden produir resultats distorsionats si estan afectats per biaixos.

Lectura recomanada

M. Ananny; K. Crawford (2016). «Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability» [en línia]. *New Media & Society*. <doi.org/10.1177/1461444816676645>

Finalment, tant pel que fa als algorismes com a les dades, sorgeix el problema de l'ús descontextualitzat d'aquestes eines i recursos. Per això, no és rar que les formes intensives d'anàlisis de dades se centrin en informació descontextualitzada, sense referències al context d'origen útils tant per a comprendre plenament els resultats del tractament com per a comprendre les seves utilitats en termes de l'aplicació concreta. De manera anàloga, els models algorítmics originalment creats per a l'anàlisi d'un problema donat es poden utilitzar posteriorment per a diferents propòsits en diferents contextos. O bé, models elaborats sobre la base de dades històriques d'una població es poden usar pel que fa a una població diferent, sense que s'hagi verificat acuradament l'efectivitat inalterada dels mateixos models en el cas d'ús descontextualitzat.

Informació descontextualitzada

Això pot ser el cas, per exemple, d'un sistema algorítmic que avalua el rendiment dels estudiants de diferents escoles, sense tenir en compte les variables socioeconòmiques de les famílies d'origen.

Lectura recomanada

Donovan i d'altres (2018). *Op. cit.* (pàg. 7).

3. El paper de l'anàlisi de riscos

Donades les limitacions que afecten tant l'autodeterminació de les persones en relació amb el tractament de dades com la transparència dels processos algorítmics, les normes de protecció de dades consideren cada vegada més el paper de l'avaluació de riscos com un remei preventiu per a evitar conseqüències perjudicials.

L'avaluació de riscos, en el context del processament de dades personals, no solament és una eina per a prevenir possibles perjudicis als drets i llibertats de les persones interessades, sinó que també exerceix un paper important en la dinàmica centrada en la confiança dels usuaris que sempre ha caracteritzat el desenvolupament de tecnologies. En aquest sentit, la presència d'un entorn d'IA «segur», des del punt de vista tècnic i jurídic, pot millorar la confiança dels usuaris d'aquestes aplicacions i augmentar la seva disposició a utilitzar-les.

Lectura recomanada

A. Mantelero (en premsa). «Comment to Articles 35 and 36» [en línia]. A: M. Cole; F. Boehm (dir.). *GDPR Commentary* Edward Elgar Publishing. <papers.ssrn.com/sol3/papers.cfm?abstract_id=3362747>

En aquest sentit, les preferències dels usuaris poden basar la seva confiança més correctament en una anàlisi dels riscos efectius i de les mesures que s'han adoptat per a fer front a aquests riscos que en campanyes de màrqueting o reputació de marca. Per aquesta raó, l'ús d'algorismes en el context del tractament de dades personals i l'ús creixent de tecnologies que fan un ús intensiu de dades han portat a una major atenció als possibles efectes adversos del processament de dades.

Lectures recomanades

A. Mantelero (2016). «Children online and the future EU data protection framework. Empirical evidences and legal analysis». *International Jour. Tech. Policy & Law* (vol. 2, núm. 2/3/4, pàg. 169-181).

European Data Protection Supervisor - Ethics Advisory Group, *op. cit.*

Grups d'experts i acadèmics han anat més enllà de l'esfera tradicional de protecció de dades per considerar l'impacte de l'ús de dades en els drets fonamentals i els valors socials i ètics, tant en un pla individual com col·lectiu. Si, d'una banda, aquesta extensió de l'avaluació de riscos es deu a l'efecte de les aplicacions centrades en la IA, que va molt més enllà de l'entorn del tractament de dades personals, d'altra banda, l'avaluació del respecte pels drets humans i els valors ètics i socials fa més difícil dur a terme l'avaluació d'impacte de l'ús de dades.

Lectures recomanades

Access Now (2018). *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems* (vol. 2, núm. 2/3/4, pàg. 169-181) [en línia]. <www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf>

A. Mantelero (2017). «From group privacy to collective privacy: towards a new dimension of privacy and data protection in the big data era». A: L. Taylor i d'altres (dir.). *Group Privacy New Challenges of Data Technologies*. Cham: Springer International Publishing.

A més, mentre que, en el camp de la seguretat i la gestió de la informació, els criteris i valors de referència (per exemple, la integritat de les dades) se centren en la tecnologia i, per tant, poden generalitzar-se en diversos contextos, la situació és diferent quan es prenen com a referència els valors ètics i socials. Aquests últims són, de fet, necessàriament específics i contextuals i difereixen d'una comunitat a una altra. D'això es dedueix que serà més difícil identificar un model de valor de referència per a una avaluació del risc que vagi més enllà de la protecció de dades personals i també inclogui altres drets fonamentals, com el dret a la no-discriminació, així com quan es tinguin en compte les conseqüències ètiques i socials de l'ús de la IA.

Lectura recomanada

D. Wright (2011). «A framework for the ethical impact assessment of information technology». *Ethics Inf. Technol* (núm. 13, pàg. 201).

La naturalesa contextual de l'avaluació inherent al respecte dels valors ètics i socials condueix, per tant, a reavaluar solucions operatives que permetin una major granularitat de l'anàlisi i una major proximitat a les comunitats de referència. D'aquí el creixent interès en el paper dels comitès d'experts o comitès d'ètica. Aquests comitès, que en alguns casos ja existeixen en la pràctica, haurien de determinar els valors específics que han de protegir-se en relació amb l'ús específic de les dades, proporcionant indicacions més detallades i contextuals al responsable del tractament de dades per a dur a terme una avaluació de riscos que sigui exhaustiva.

Lectura recomanada

Council of Europe. *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data, op. cit.*

En aquest sentit, ha de tenir-se en compte que el major esforç resultant d'una avaluació de riscos més àmplia no solament es deu a la naturalesa dels drets i llibertats potencialment afectats per l'aplicació de la IA i per les seves importants conseqüències socials, sinó que també representa una oportunitat per al desenvolupament d'innovació responsable, així com un avantatge competitiu potencial per a les empreses.

De fet, incrementar la confiança dels usuaris en els productes i serveis d'IA ofereix a les empreses l'oportunitat de respondre millor a les creixents preocupacions dels consumidors sobre l'ús de les seves dades i de solucions d'IA en general. De la mateixa manera, una major consideració de les conseqüències de l'adopció de sistemes d'IA per part dels organismes públics solament pot tenir un impacte positiu en termes de la confiança dels ciutadans en l'acció de l'Administració pública i en la prevenció de decisions afectades per biaixos en la presa de decisions.

Lectura recomanada

IEEE (2019). *Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems First Edition Overview* [en línia]. <ethicsinaction.ieee.org>

Certificacions, codis de conducta i estàndards

En aquesta perspectiva, cal destacar que les certificacions, els codis de conducta i els estàndards també poden exercir un paper important. Aquestes eines diferents ajuden en la rendició de comptes i proporcionen orientació sobre la integritat de les dades i del sistema.

Finalment, cal destacar que l'avaluació de les conseqüències de l'ús dels sistemes d'IA, donat l'impacte que solen tenir en una pluralitat de persones i en comunitats senceres, no ha de dur-se a terme sense la participació acti-

[5] Vegeu l'art. 35, par. 9, Regl. 2016/679.

va d'aquestes comunitats. Aquest és el tema de l'enfocament participatiu en l'avaluació de riscos, ja considerat pel legislador europeu pel que fa al processament de dades personals,[5] però encara més rellevant si volem crear un model més ampli d'anàlisi de riscos que inclogui les conseqüències socials.

Lectures recomanades

Article 29 Data Protection Working Party, *op. cit.*

Un enfocament participatiu també pot ser útil per a una millor comprensió dels diversos interessos i valors en joc, pel que fa a aplicacions específiques d'IA. En aquest sentit, la participació de les persones interessades també representa un objectiu de l'avaluació d'impacte, ja que contribueix a reduir el risc de subrepresentació d'alguns grups o categories de persones i també és potencialment capaç de destacar aspectes crítics que se subestimen o ignoren en l'avaluació abstracta realitzada pel responsable del tractament de dades.

Lectura recomanada

United Nations Office of the High Commissioner for Human Rights (2016). *Frequently asked questions on a human rights-based approach to development cooperation*. Nova York / Gènova: United Nations.

Òbviament, la participació de les parts interessades no ha de veure's com una forma en què els qui prenen les decisions (els responsables del tractament de dades en aquest cas) poden eludir les seves responsabilitats com a subjectes que han de gestionar tot el procés. Més aviat, el propòsit d'una avaluació participativa dels efectes de llarg abast de la presa de decisions algorítmiques és induir els responsables del tractament a adoptar solucions de codisseny per al desenvolupament d'aplicacions d'IA, per tal d'involucrar activament els grups potencialment afectats.

Resum

Amb aquests temes de reflexió que s'han tractat a les pàgines anteriors, es volia dur a terme una anàlisi dels problemes crítics que planteja el recurs creixent a la IA pel que fa a la protecció de dades personals i, en general, pel que fa als drets i les llibertats de les persones, així com a les possibles conseqüències socials.

Es tracta d'un tema de recerca que és en bona part nou i obert, que s'està allunyant de les formes de protecció i mesures dissenyades exclusivament per a la protecció de dades personals, que semblen cada vegada més limitades per a abordar de manera adequada els aspectes més crítics de la IA i de les conseqüències que comporta la societat algorítmica. Una prova d'aquest esforç per mirar més enllà són les diverses iniciatives desenvolupades a escala internacional pel Consell d'Europa, l'EDPS, el Parlament Europeu, la Comissió Europea i moltes altres entitats, amb la finalitat de dirigir la nostra mirada cap a un escenari més ampli de criteris i valors sobre els quals construir un futur marc regulatori. Un marc que també inclou les conseqüències socials de l'ús de la IA i que avalua en un sentit ampli, sense limitar-se només al dret de les dades personals, l'impacte de les aplicacions d'aquesta tecnologia en els drets i les llibertats de les persones.

Tot i que encara queda molt de camí per recórrer abans que es puguin establir models comuns i consolidats, el debat científic sobre aquests temes i les iniciatives esmentades abans ja han marcat la direcció principal.

Bibliografia

Access Now (2018). *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems* [en línia]. <https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf>

ACM (2018). *ACM Code of Ethics and Professional Conduct* [en línia]. <https://www.acm.org/code-of-ethics>

AI Now Institute (2016). *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term* [en línia]. <https://ainowinstitute.org/AI_Now_2016_Report.pdf>

AI Now Institute (2017). *AI Now 2017 Report* [en línia]. <https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf>

AI Now Institute (2018). *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems* [en línia]. <https://ainowinstitute.org/litigatingalgorithms.pdf>

Ananny, M.; Crawford, K. (2016). «Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability» [en línia]. *New Media & Society*. doi: <https://doi.org/10.1177/1461444816676645>

Artificial Intelligence Index. Annual Report 2017 [en línia]. <http://aiindex.org/2017-report.pdf>

Asilomar AI Principles 2017 [en línia]. <https://futureoflife.org/ai-principles/>

Axon AI and Policing Technology Ethics Board [en línia]. <https://www.axon.com/axon-ai-and-policing-technology-ethics>

Barocas, S.; Nissenbaum, H. (2015). «Big Data's End Run around Anonymity and Consent». A: L. Lane; V. Stodden; S. Bender; H. Nissenbaum (ed.). *Privacy, big data, and the public good: frameworks for engagement*. Cambridge: Cambridge University Press.

Barocas, S.; Selbstr, A. D. (2016). «Big Data's Disparate Impact». *California Law Review* (vol. 104, núm. 3, pàg. 671-732).

Barse, E. L.; Kvarnstrom, H.; Jonsson, E. (2003). «Synthesizing Test Data for Fraud Detection Systems». A: *19th Annual Computer Security Applications Conference. Proceedings* (pàg. 384-394). doi: <https://doi.org/10.1109/CSAC.2003.1254343>

Binns, R. i d'altres (2018). «It's Reducing a Human Being to a Percentage» [en línia]. *Perceptions of Justice in Algorithmic Decisions*. ArXiv:1801.10408 [Cs], (pàg. 1-14). doi: <https://doi.org/10.1145/3173574.3173951>

Bostrom, N. (2016). *Superintelligence paths, dangers, strategies*. Oxford: Oxford University Press.

Boyd, D.; Crawford, K. (2012). «Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon». *Information, Communication, & Society* (vol. 15, núm. 5, pàg. 662-679).

Brauneis, R.; Goodman, E. P. (2018). «Algorithmic Transparency for the Smart City». *Yale J. L. & Tech* (vol. 20, pàg. 103-176).

Bray, P. i d'altres (2015). *International differences in ethical standards and in the interpretation of legal frameworks SATORI Deliverable D3.2* [en línia]. <http://satoriproject.eu/work_packages/legal-aspects-and-impacts-of-globalization/>

Brundage, M. i d'altres (2018). «The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation» [en línia]. <https://maliciousaireport.com/>

Burrell, J. (2016). «How the machine "thinks": Understanding opacity in machine learning algorithms». *Big Data & Society* (vol. 3, núm. 1). doi: <https://doi.org/10.1177/2053951715622512>

Burt, A.; Leong, B.; Shirrell, S. (2018). «Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models». *Future of Privacy Forum*.

Bygrave L. A. (2001). «Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling». *Computer Law & Security Rev.* (vol. 17, núm. 1).

Calo, R. (2013). «Consumer Subject Review Boards: A Thought Experiment» [en línia]. *Standford Law Review* (vol. 66). <http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-review-boards>

Caruana, R.; Lou, Y.; Gehrke, J.; Koch, P.; Sturm, M.; Elhadad, N. (2015). «Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission». A: *Proceedings of the 21st Annual SIGKDD International Conference on Knowledge Discovery and Data Mining* (pàg. 1721-1730).

Citron, D. K.; Pasquale, F. (2014). «The Scored Society: Due Process For Automated Predictions». *Washington Law Review* (vol. 89, pàg. 1-33).

CNIL - LINC (2017). *La Plateforme d'une Ville Les Données Personnelles Au Coeur de La Fabrique de La Smart City* [en línia]. <https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf>

CNIL (2017). *How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence* [en línia]. Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill (pàg. 14). <https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf>

Conseil National du Numérique (2015). *Ambition numérique: pour une politique francaise et europeéenne de la transition numérique* [en línia]. <https://cnnumerique.fr/nos-travaux/ambition-numerique>

Council of Europe (2017). *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data* [en línia]. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a>

Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (2018). *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications* [en línia]. <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>

Cummings, M.; Roff, L.; Heather, M.; Cukier, K.; Parakilas, J.; Bryce, H. (2018). *Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated* [en línia]. Londres: Chatham House. The Royal Institute of International Affairs. <https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf>

Diakopoulos, N. (2013). *Algorithmic Accountability Reporting: on the Investigation of Black Boxes (Tow Center for Digital Journalism).*

DNA Web Team (2018, abril). «Google drafting ethical guidelines to guide use of tech after employees protest defence project» [en línia]. *DNA India*. <http://www.dnaindia.com/technology/report-google-drafting-ethical-guidelines-to-guide-use-of-tech-after-employees-protest-defence-project-2605149>

Donovan, J.; Matthews, J.; Caplan, R.; Hanson, L. (2018, abril). *Algorithmic Accountability: A Primer* [en línia]. <https://datasociety.net/output/algorithmic-accountability-a-primer/>

Doshi-Velez, F. i d'altres (2017). *Accountability of AI Under the Law: The Role of Explanation* [en línia]. <https://cyber.harvard.edu/publications/2017/11/AIExplanation>

Edwards, L.; Vale, M. (2017). «Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For». *Duke Law and Technology Review* (vol. 16, núm. 1, pàg. 18-84).

European Commission (2018). *The European Artificial Intelligence Landscape* [en línia]. <https://ec.europa.eu/digital-single-market/en/news/european-artificial-intelligence-landscape>

European Commission - European Group on, Ethics in Science and, & New Technologies (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems* [en línia]. <https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf>

European Data Protection Supervisor (2016). *Opinion 8/2016. EDPS Opinion on coherent enforcement of fundamental rights in the age of big data*.

European Data Protection Supervisor - Ethics Advisory Group (2018). *Towards a digital ethics* [en línia]. <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf>

European Economic and Social Committee (2017). *The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context* [en línia]. <https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data>

European Parliament (2017). *European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI))* [en línia]. <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//EN&language=EN>

European Union Agency for Fundamental Rights (FRA) (2018). *#BigData: Discrimination in Data-Supported Decision Making* [en línia]. <http://fra.europa.eu/en/publication/2018/big-data-discrimination>

Executive Office of the President, and National Science and Technology Council - Committee on Technology (2016). *Preparing for the Future of Artificial Intelligence* [en línia]. Washington D.C. <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf>

Federal Ministry of Transport and Digital Infrastructure (2017). *Ethics Commission Automated and Connected Driving* [en línia]. <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile>

Gama, J. i d'altres (2013). «A survey on concept drift adaptation» [en línia]. *ACM Computing Surveys* (vol. 1, núm. 1). <http://www.win.tue.nl/~mpechen/publications/pubs/Gama_ACMCS_AdaptationCD_accepted.pdf>

Goodman, B.; Flaxman, S. (2016). «EU Regulations on Algorithmic Decision-Making and a "right to Explanation"» [en línia]. arXiv:1606.08813 [cs, stat]. <http://arxiv.org/abs/1606.08813>

Hildebrandt, M. (2016). *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Cheltenham: Edward Elgar Publishing.

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (2016). *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems* (Version 1) [en línia]. IEEE. <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

Information Commissioner's Office (2017). *Big Data, Artificial Intelligence, Machine Learning and Data Protection* [en línia]. <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

ITU (2017). *AI for Good Global Summit Report 2017* [en línia]. <https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf>

Kaye, J. i d'altres (2015). «Dynamic consent: a patient interface for twenty-first century research networks» [en línia]. *European Journal of Human Genetics* (vol. 23, núm. 2, pàg. 141). <https://www.nature.com/articles/ejhg201471>

Kurzweil, R. (2016). *The singularity is near: when humans transcend biology*. Londres: Duckworth.

Linnet, T.; Floridi, L.; van der Sloot, B. (ed.). (2017). *Group Privacy: New Challenges of Data Technologies*. Nova York: Springer International Publishing.

Lipton, Z. C. (2018). «The Mythos of Model Interpretability. In Machine Learning, the Concept of Interpretability Is Both Important and Slippery» [en línia]. *ACMQueue* (vol. 16, núm. 3). <https://queue.acm.org/detail.cfm?id=3241340>

Lomas, N. (2017). «DeepMind now has an AI ethics research unit. We have a few questions for it...» [en línia]. *TechCrunch*. <http://social.techcrunch.com/2017/10/04/deepmind-now-has-an-ai-ethics-research-unit-we-have-a-few-questions-for-it/>

Lycett, M. (2013). «Datafication: making sense of (big) data in a complex world». *European Journal of Information Systems* (vol. 22, núm. 4, pàg. 381-386).

Mantelero, A. (2014). «The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics». *Computer Law and Security Review* (vol. 30, núm. 6, pàg. 643-660).

Mantelero, A. (2017). «Regulating Big Data. The guidelines of the Council of Europe in the Context of the European Data Protection Framework». *Computer Law & Sec. Rev.* (vol. 33, núm. 5, pàg. 584-602).

Mantelero, A. (2018). «AI and Big Data: A blueprint for a human rights, social and ethical impact assessment» [en línia]. *Assessment. Computer Law & Security Review*. doi: <https://doi.org/10.1016/j.clsr.2018.05.017>

Mayer-Schönberger, V.; Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think*. Londres: John Murray.

McCulloch, W. S.; Pitts, W. H. (1943). «A Logical Calculus of the Ideas Immanent in Nervous Activity». *Bulletin of Mathematical Biophysics* (vol. 5, pàg. 115-133).

Office of the Privacy Commissioner of Canada (2016). *The Internet of Things: An Introduction to Privacy Issues with a Focus on the Retail and Home Environments* [en línia]. <https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/iot_201602/#heading-0-0-2-15>

O'Neil, C. (2017). *Weapons of math destruction*. Londres: Penguin Books.

Palm, E.; Hansson, S. O. (2006). «The case for ethical technology assessment (eTA)». *Technological Forecasting & Social Change* (vol. 73, núm. 5, pàg. 543, 550-551).

Polonetsky, J.; Tene, O.; Jerome, J. (2015). «Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings». *Colorado Technology Law Journal* (vol. 13, pàg. 333-367).

Raso, F. i d'altres (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks* [en línia]. <https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf?subscribe=Download+the+Report>

Reisman, D.; Schultz, J.; Crawford, K.; Whittaker, M. (2018). *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* [en línia]. <https://ainowinstitute.org/aiareport2018.pdf>

Rossi, F. (2016). *Artificial Intelligence: Potential Benefits d Ethical Considerations' (European Parliament: Policy Department C: Citizens' Rights and Constitutional Affairs 2016)* [en línia]. Briefing PE 571.380. <http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf>

Rouvroy, A. (2016). *«Of Data and Men»: Fundamental Rights and Liberties in a World of Big Data* [en línia]. Estrasburg: Consell d'Europa. <https://pure.unamur.be/ws/portalfiles/portal/13278394/Report_Big_Data.pdf>

Rubinstein, I. S. (2013). «Big Data: The End of Privacy or a New Beginning?». *International Data Privacy Law* (vol. 3, núm. 2, pàg. 74-87).

Schreurs W.; Hildebrandt M.; Kindt E.; Vanfleteren M. (2010). «Cogitas, Ergo Sum. The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector». A: M. Hildebrandt-S; Gutwirth (ed.). *Profiling the European Citizen. Cross-Disciplinary Perspective* (pàg. 241). Dordrecht: Springer.

Selbst, A. D. (2017). «Disparate Impact in Big Data Policing». *Georgia Law Review* (vol. 52, núm. 1, pàg. 109-195).

Selbst, A. D.; Powles, J. (2017). «Meaningful Information and the Right to Explanation». *International Data Privacy Law* (vol. 7, núm. 4, pàg. 233-242).

Sheehan, M. (2011). «Can Broad Consent be Informed Consent?». *Public Health Ethics* (vol. 3, pàg. 226-235).

Spiekermann, S. (2016). *Ethical IT Innovation. A Value-Based System Design Approach*. Boca Raton: CRC Press.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. (2013). *Intriguing properties of neural networks* [en línia]. <https://arxiv.org/abs/1312.6199>

Tene, O.; Polonetsky, J. (2012). «Privacy in the Age of Big Data. A Time for Big Decisions». *Stanford Law Review Online* (vol. 64 pàg. 63-69).

The Danish Institute for Human Rights (2016). *Human rights impact assessment guidance and toolbox* [en línia]. <https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-and-toolbox>

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (2016). *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems* (Version 1) [en línia]. <http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html>

The Norwegian Data Protection Authority (2018). *Artificial Intelligence and Privacy Report* [en línia]. <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>

Turing, A. M. (1950). «Computing Machinery and Intelligence». *Mind* (vol. 49, pàg. 433-460).

UK Department for Digital, Culture, Media & Sport. *Data Ethics Framework - GOV.UK* [en línia]. <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework>

United Nations Office of the High Commissioner for Human Rights (2006). *Frequently asked questions on a human rights-based approach to development cooperation*. Nova York / Gènova: United Nations.

United Nations (2011). *Guiding Principles on Business and Human Rights: Implementing the United Nations «Protect, Respect and Remedy» Framework*. United Nations Human Rights Council (UN Doc. HR/PUB/11/04).

Veale M.; Binns R. (2017). «Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data» [en línia]. *Big Data & Society* (vol. 4, núm. 2, 2053951717743530). doi: <https://doi.org/10.1177/2053951717743530>

Veale M.; Edwards L. (2018). «Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling». *Computer Law & Security Review* (vol. 34, núm. 2, pàg. 398-404).

Veale, M.; Binns, R.; Edwards, L. (2018). «Algorithms That Remember: Model Inversion Attacks and Data Protection Law» [en línia]. *Philosophical Transactions of the Royal Society*. doi: <https://doi.org/10.1098/rsta.2018.0083>

Villani, C. (2018). *For a Meaningful Artificial Intelligence towards a French and European Strategy* [en línia]. <https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf>

Wachter, S.; Mittelstadt, B.; Loridi, L. (2017). «Why a right to explanation of automated decision - making does not exist in the General Data Protection Regulation» [en línia]. *International Data Privacy Law* (vol. 7, núm. 2, pàg. 76-99). <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469>

Walker, S. M. (2009). *The Future of Human Rights Impact Assessments of Trade Agreements* [en línia]. Utrecht: G. J. Wiarda Institute for Legal Research. <https://dspace.library.uu.nl/bitstream/handle/1874/36620/walker.pdf?sequence=2>

White House (2015). *Consumer Privacy Bill of Rights Act*. §103(c). Administration Discussion Draft 2015. <https://obamawhitehouse.archives.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf>

World Economic Forum (2018). *How to Prevent Discriminatory Outcomes in Machine Learning* [en línia]. <http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf>

Wright, D. (2011). «A framework for the ethical impact assessment of information technology». *Ethics and Information Technology* (vol. 13, pàg. 199, 201-202).

Wright, D.; De Hert, P. (ed). (2012). *Privacy Impact Assessment*. Dordrecht: Springer.

Wright, D.; Mordini, E. (2012). «Privacy and Ethical Impact Assessment». A: D. Wright; P. De Hert (ed.). *Privacy Impact Assessment* (pàg. 397-418). Dordrecht: Springer.