Fordham Intellectual Property, Media and Entertainment Law Journal

Volume 30 *XXX*
Number 1

Article 3

2019

Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making

Céline Castets-Renard
University of Ottawa, Celine.Castets-Renard@uottawa.ca

Recommended Citation

Céline Castets-Renard, *Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making*, 30 Fordham Intell. Prop. Media & Ent. L.J. 91 (2019).
Available at: https://ir.lawnet.fordham.edu/iplj/vol30/iss1/3

Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making

Cover Page Footnote

Professor, University of Ottawa, Civil Law Faculty (Canada), Chair of the Law, Accountability, Social Trust in AI, ANITI (Artificial and Natural Intelligence Toulouse Institute) (France), and Adjunct Professor, Fordham University School of Law (New York). This Article has been presented to a workshop at the Fordham Center on Law and Information Policy (CLIP) in April 2018. The author thanks the participants for their useful comments: Marie-Apolline Barbara, Visiting Researcher, Cornell Tech; Erin Carroll, Associate Professor of Legal Research and Writing, Georgetown University; Tithi Chattopadhyay, Associate Director, Center for Information Technology Policy (CITP), Princeton University; Ignacio Cofone, J.S.D. candidate at Yale Law School and Research Fellow at NYU, Information Law Institute; Roger A. Ford, Associate Professor of Law, University of New Hampshire School of Law; Frank Pasquale, Professor of Law, University of Maryland Carey School of Law; N. Cameron Russell, Executive Director of CLIP, Adjunct Professor of Law, Fordham University School of Law. Thanks to the Fulbright-French Commission Program for the Fulbright Grant and to Professor Joel Reidenberg, Stanley D. and Nikki Waxberg Chair and Professor of Law, Founding Academic Director, Fordham CLIP for welcoming me. I also thank Professor Jack Balkin, Knight Professor of Constitutional Law and the First Amendment at Yale Law School, and founder and director of Yale’s Information Society Project (ISP), as well as Rebecca Crootof, Executive Director of the ISP, Research Scholar and Lecturer in Law at Yale Law School for welcoming me as ISP Visiting Fellow.

Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making

Céline Castets-Renard*

Automated decision systems appear to carry higher risks today than they ever have before. Digital technologies collect massive amounts of data and evaluate people in every aspect of their lives, such as housing and employment. This collected information is ranked through the use of algorithms. The use of such algorithms may be problematic. Because the results obtained through

[*] Professor, University of Ottawa, Civil Law Faculty (Canada), Chair of the Law, Accountability, Social Trust in AI, ANITI (Artificial and Natural Intelligence Toulouse Institute) (France), and Adjunct Professor, Fordham University School of Law (New York). This Article has been presented to a workshop at the Fordham Center on Law and Information Policy (CLIP) in April 2018.

The author thanks the participants for their useful comments: Marie-Apolline Barbara, Visiting Researcher, Cornell Tech; Erin Carroll, Associate Professor of Legal Research and Writing, Georgetown University; Tithi Chattopadhyay, Associate Director, Center for Information Technology Policy (CITP), Princeton University; Ignacio Cofone, J.S.D. candidate at Yale Law School and Research Fellow at NYU, Information Law Institute; Roger A. Ford, Associate Professor of Law, University of New Hampshire School of Law; Frank Pasquale, Professor of Law, University of Maryland Carey School of Law; N. Cameron Russell, Executive Director of CLIP, Adjunct Professor of Law, Fordham University School of Law. Thanks to the Fulbright-French Commission Program for the Fulbright Grant and to Professor Joel Reidenberg, Stanley D. and Nikki Waxberg Chair and Professor of Law, Founding Academic Director, Fordham CLIP for welcoming me. I also thank Professor Jack Balkin, Knight Professor of Constitutional Law and the First Amendment at Yale Law School, and founder and director of Yale's Information Society Project (ISP), as well as Rebecca Crootof, Executive Director of the ISP, Research Scholar and Lecturer in Law at Yale Law School for welcoming me as ISP Visiting Fellow.

algorithms are created by machines, they are often assumed to be immune from human biases. However, algorithms are the product of human thinking and, as such, can perpetuate existing stereotypes and social segregation. This problem is exacerbated by the fact that algorithms are not accountable. This Article explores problems related to algorithmic bias, error, and discrimination which exists due to a lack of transparency and understanding behind a machine's design or instruction.

This Article deals with the European Union's legal framework on decision-making on the General Data Protection Regulation ("GDPR") and some Member State implementation laws, with specific emphasis on French law. This Article argues that the European framework does not adequately address the algorithm's problems of opacity and discrimination related to machine learning processing and the explanations of automated decision-making. The Article proceeds by evaluating limitations to the legal remedies provided by the GDPR. In particular, the GDPR's lack of a right to individual explanation regarding these decisions poses a problem. Furthermore, the Article also argues that the GDPR allows for too many flexibilities for individual Member States, thus failing to create a "digital single market." Finally, this Article proposes certain solutions to address the opacity and bias problems of automated decision-making.

INTRODUCTION

Today, automated decision systems appear to carry higher social and economic risks than ever before. We often have no information about the design or instructions the machine is given. This easily becomes a source of biases, errors, and discrimination. Indeed, an algorithm is not neutral and can perpetuate existing stereotypes and social segregation. For example, underrepresentation of a minority group in historical data may reinforce discrimination against that group in future hiring processes or credit-scoring.

This Article's subject matter deals with the European Union's ("EU") legal framework on automated decision-making based on the GDPR and some Member State implementation laws with specific emphasis on French law. In Part I, I discuss the current role automated decision-making plays in our society and the need for more ethics and rulemaking to eliminate opacity and bias problems in such technology. In Part II, I present the European legal framework. Currently, the European Union and its Member States have enacted a more precise framework on automated decision-making, based on the GDPR on civil and commercial matters as well as on the Directive 2016/680/EU on criminal matters. The GDPR is completed by guidelines from the Article 29 Working Party.[1] However, I argue in particular that there is no right to an individual explanation concerning a decision based on automated decision-making pursuant to the GDPR. The GDPR does not provide the data subject with an individual right to know and understand the automated decision's precise basis.

In Part III, I argue that, if EU lawmakers understand the issues, their answers are not strong enough to improve the rules and protect the vulnerable population. The exceptions give too many flexibilities in favor of private stakeholders, public sectors, and the Member States. Compounding the exceptions, the related safeguards, such as the right to obtain a human intervention, do not provide for a right to an explanation either; they only afford the right to ask for a human being, and not a machine, with whom to interact. Nevertheless, this right does not ensure a better understanding of the decision. Indeed, it may not be feasible for a human to conduct a meaningful review of a process—for instance, if the process involved third-party data and algorithms, pre-learned models, or inherently opaque machine learning techniques. Moreover, intellectual property rights and trade secrets create some barriers to

[1] Article 29 Working Party (Art. 29 WP**)** was an advisory body made up of a representative from the data protection authority of each EU Member State, the European Data Protection Supervisor and the European Commission. On May 25, 2018, it was replaced by the European Data Protection Board (EDPB) under the GDPR. All documents related to the former Article 29 Working Party remain available, such as non-binding guidelines which provide interpretations of EU directives and regulations on data protection laws.

the rights' efficiencies, and the GDPR does not furnish limitations to the application of such proprietary rights in the privacy context. Finally, no supervisory body explicitly provides for guarantees to respect such measures. Consequently, I am skeptical as to the ability of such provisions to address the opacity and discrimination problems of algorithms.

I also argue that too many flexibilities have been given to the Member States, creating a variety of differing rules. After the integration of the "EU Personal Data Package" at the national level, one can see that the common rules between the Member States are less numerous than expected. Consequently, despite the enactment of an EU Regulation instead of a Directive,[2] the European rules are too weak and too diverse to adequately protect Europeans. As a result, the GDPR also fails to create a single standard on algorithmic transparency. This has a negative impact on the ability to create a "digital single market," which is one of the European Commission's primary goals.[3]

Finally, in Part IV, I consider what might be done to formulate a better framework. I propose some solutions, which have to be challenged and improved.

I. OPACITY, BIAS PROBLEMS OF AUTOMATED DECISION-MAKING, AND THE NEED FOR MORE ETHICS

A. Effects of "Automated Decision-Making"

Today we live in a "Scored Society"[4] or "Black Box Society."[5] Digital technologies collect massive data (big data) and score people

[2] Regulations have binding legal force throughout every Member State and enter into force on a set date in all the Member States. Directives lay down certain results that must be achieved but each Member State *is* free to decide how to transpose the directives into national laws.

[3] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe*, at 3, COM (2015) 192 final (May 6, 2015).

[4] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 2 (2014).

[5] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 10 (2015).

in every aspect of their lives: what they buy, what they do, what they think, how they work, what their networks are, and how they conduct their personal and intimate lives.[6] Any and all information can be collected and coded to produce an opinion on an individual or to provide a right to access to an advantage, or the denial of such. The information that is collected can be used to generate rankings used in many circumstances, such as in job applications, social benefits, or loans.[7] A person's online activity, like their interactions with social networks, is an example of the kind of information used to generate rankings.[8] This scoring system is made by algorithms instead of humans.[9] As a result of predictive algorithms making essential decisions about individuals, one's personal life can change.[10] More broadly, this means that economic activities change: financial markets, marketing, insurance, employment, education, political elections, judicial decisions, and so on. Many scholars have already shown the effects of predictive algorithms on both individual and collective situations.[11]

Basically, an "algorithm" is a sequence of instructions telling a computer what to do. In the broadest sense, algorithms "are encoded procedures for transforming input data into a desired output, based on specified calculations."[12] This notion is broad and includes "artificial intelligence" processing, which itself contains machine learning and deep learning.[13] The term "artificial intelligence" applies when a machine mimics "cognitive" functions associated with human minds, such as "learning" and "problem-solving."[14] "Machine learning" is supposed to give a computer system the

[6] *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 15–17 (2013).

[7] *See* Citron & Pasquale, *supra* note 4, at 2, 5, 28.

[8] *See id.* at 2.

[9] *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 638, 658 (2017).

[10] *See* Citron & Pasquale, *supra* note 4, at 20.

[11] *See, e.g.*, CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 3, 13, 199–204 (2016).

[12] Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 167 (Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot eds., 2014).

[13] *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 90 (2014).

[14] *Id.* at 89, 94.

ability to progressively improve performance on a specific task, based on the use of data mining and massive gathering ("big data"), without explicit programming.[15] This is what is known as unsupervised learning.[16] Machine learning methods are based on learning data representations as opposed to task-specific algorithms.[17] This means that the machine "learns" by itself, in consideration of a goal previously defined by the programmer. In contrast, human intervention is mainly focused on the definition of goals (task-specific algorithms) and data used.[18] These tools analyze current and historical facts, allowing the models to make predictions (predictive models). Finally, "deep learning" architectures, such as deep neural networks, have been applied to fields including computer vision, speech recognition, and natural language processing.[19] Automated individual decision-making is based more on machine learning than on deep learning.

Though algorithms may be problematic in some ways, several positive elements exist as well. First, an automated decision-making process may be more efficient than its alternative: the information gained can be more useful and cheaper to obtain than the information gathered through human decision-making.[20] Second, although the possibility exists that an algorithm is biased, such bias often occurs because automated decision-makers were trained using biased human decisions.[21] Of course, not all training data is based on bias. One such example is credit scoring, which is based on actual payment data, not human assessments of creditworthiness.[22] This is

[15] *See* Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data*, BIG DATA & SOC'Y, July–Dec. 2017, at 1, 11.

[16] *Id.*

[17] *See* Yann LeCun, Yoshua Bengio & Geoffrey Hinton*, Deep Learning*, 521 NATURE 436, 436, 442 (2015).

[18] *See id.* at 442.

[19] *See* DAN CIREȘAN, UELI MEIER & JÜRGEN SCHMIDHUBER*,* DALLE MOLLE INST. ARTIFICIAL INTELLIGENCE, MULTI-COLUMN DEEP NEURAL NETWORKS FOR IMAGE CLASSIFICATION 1 (2012); Surden, *supra* note 13, at 99.

[20] *See* Jay Thornton, *Cost, Accuracy, and Subjective Fairness in Legal Information Technology: A Response to Technological Due Process Critics*, 91 N.Y.U. L. REV. 1821, 1822 (2016).

[21] *See* Citron & Pasquale, *supra* note 4, at 4.

[22] *See id.* at 5.

an example of a situation where an algorithm may actually be less biased than human judgment. Human bias is most likely to exist in an algorithm where training data has been filtered through human intervention.[23] This is why sentencing and arrest data are bad training data.

Third, algorithms are usually more accurate than the alternative.[24] Humans make messy, error-ridden assessments of multi-dimensional information in decision-making and are subject to numerous cognitive biases. Algorithms seem to make less error-prone assessments and seem to be less subject to biases.[25]

Nevertheless, we must not forget the social need to better understand algorithms and their resulting decisions.[26] Humans must maintain control of, and be accountable for, the decisions made by machines. For example, the scoring process is often seen as a good method.[27] It is considered progress in society because it is supposed to be more objective and non-discriminatory than human decision-making.[28] However, this is a common mistake. Algorithms are not neutral and can perpetuate existing stereotypes and social segregation.[29] Additionally, big data analytics, artificial intelligence, and machine learning's capabilities have significantly facilitated the creation of profiles and automated decisions with the potential to impact individual's rights and freedoms—especially when the decision concerns an application to enter a school or to obtain social benefits.[30]

[23] *See id.* at 4.

[24] *See* Thornton, *supra* note 20, at 1825.

[25] *See id.* at 1835.

[26] *See* Mireille Hildebrandt, *The Dawn of a Critical Transparency Right for the Profiling Era*, *in* DIGITAL ENLIGHTENMENT YEARBOOK 2012, at 42 (Jacques Bus et al. eds., 2012).

[27] *See, e.g.*, Citron & Pasquale, *supra* note 4, at 4.

[28] *Id.*

[29] *See* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1030–31, 1038 (2017).

[30] *See, e.g.*, Germain Chastel, *Predictive Algorithms Are Infiltrating Schools—Here's Why That's a Good Thing*, NEXT WEB (May 28, 2018), https://thenextweb.com/contributors/2018/05/27/predictive-algorithms-are-infiltrating-schools-heres-why-thats-a-good-thing/ [https://perma.cc/B2NK-FY5S]; Jacob Mchangama & Hin-Yan Liu, *The Welfare State Is Committing Suicide by Artificial Intelligence*, FOREIGN POL'Y (Dec. 25, 2018, 1:00 AM), https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/ [https://perma.cc/K59K-NM3U].

Several arguments show the limits of algorithmic decision-making. First, predictive algorithms are based on source code, meaning that some instructions have been given and some data has been used. A bias problem exists when a computer system systematically and unfairly discriminates against groups of individuals whilst favoring others based on social or ethical criteria.[31] AI-based technologies are developed by people who may hold explicit or implicit biases against members of underrepresented groups. Bias may be introduced into machine learning processes at various stages, including algorithm design.[32] Most often, we have no information about the design or instructions given to the machine, and these could easily be a source of biases, errors, and discrimination.

Second, bias can also be implicit,[33] as some of the processes by which the brain uses mental associations are so well-established as to operate without awareness, intention, or control (e.g., the "White Guy problem").[34] "Preexisting bias has its roots in social institutions, practices, and attitudes."[35] We usually have no information on the nature and source of data,[36] and many AI systems learn to make classifications by training on data sets that reflect sociocultural biases.[37] It is unsurprising that outputs of technologies replicate inequalities when they have been taught using biased data.[38] Selection of training data may embed existing prejudices into automated decision-making processes. For example, under-

[31] *See* Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS. 330, 332 (1996).

[32] *See* WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 14–15, 43, 51, 134 (2018).

[33] SARAH E. REDFIELD, ENHANCING JUSTICE: REDUCING BIAS 1 (Sarah E. Redfield ed., 2017).

[34] Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/ZMW2-5ABH] ("[A]lgorithmic flaws aren't easily discoverable: How would a woman know to apply for a job she never saw advertised? How might a black community learn that it were being overpoliced by software?").

[35] Friedman & Nissenbaum, *supra* note 31, at 332.

[36] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 684, 717–18 (2016).

[37] *See* Friedman & Nissenbaum, *supra* note 31, at 333.

[38] *See* Citron & Pasquale, *supra* note 4, at 13–14, 18.

representation of a minority group in historical data may reinforce discrimination against that group in future hiring processes or credit-scoring.[39] "Profiling based on postal codes or even magazine subscriptions may become a proxy for selection based on race or gender."[40]

Beyond the bias problem, the opacity of the models must also be considered. Opacity results in a reduced margin of error while rendering interpretation, human explanation, and recommendation impossible.[41] As the machine "learns" by itself, and human intervention is mainly focused on the definition of task-specific algorithms and data used, humans are not able to explain the decision-making. Furthermore, the reasoning of the machine (artificial intelligence) is not comparable to natural intelligence. The machine does not "think" as a human. Consequently, a human being is not able to pursue the lines of thinking the machine employs, and the results produced cannot be transparent and explainable. Human understanding is sacrificed in favor of an engineering perspective. This is the "black box," meaning that we do not understand the results and decisions made by algorithms.[42] Data scientists increasingly cannot explain the processes through which algorithms operate; they only find the efficiency of the results. Moreover, correlations and inferences replace causality. Consequently, these technical and legal obstacles establish asymmetric information between, on the one hand, the users of the algorithm system and, on the other hand, the persons about whom the results are generated. In such circumstances, the results cannot be audited, which is probably the best way to become aware of bias and discrimination problems.

Confirming these criticisms, Jeff Larson and his coauthors[43] denounced bias of the predictive justice system, COMPAS,[44] which

[39] *See, e.g.*, Barocas & Selbst, *supra* note 36, at 684–85.

[40] Christopher Kuner et al., *Machine Learning with Personal Data: Is Data Protection Law Smart Enough to Meet the Challenge?*, 7 INT'L DATA PRIVACY L. 1, 2 (2017).

[41] *See* Citron & Pasquale, *supra* note 4, at 10–11, 31.

[42] *See id.* at 6, 8.

[43] Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [https://perma.cc/5CYH-GEJR].

[44] COMPAS stands for Correctional Offender Management Profiles for Alternative Sanction and was created by Northpointe Inc. *See* Tim Brennan & William Dietrich,

is designed to predict the likelihood of recidivism.[45] This system is used by several states in the United States at sequential stages of criminal justice, including at pretrial and community corrections, probation, jail, prison, and parole.[46] Its goals include accurate risk assessment, comprehensive needs assessment, public safety, institutional safety, fairness and racial equity, and ease of use.[47] The risk of discrimination implicates the protection of citizens' fundamental rights, and Larson and others show that the rate of false positives (high score of risk without observed recidivism) is more frequent for Afro-American released prisoners than for Caucasian released prisoners.[48] Alexandra Chouldechova has shown that the learning sample, rather than the model, is biased, because the sample reflects preexisting social biases.[49] She then proves how disparate impact can arise when a recidivism prediction instrument fails to satisfy the criterion of error rate balance.[50] Consequently, even though COMPAS pretends to conduct periodic re-validation, re-forming, and calibration studies,[51] there is a risk of increasing these biases.[52] Despite the evidence of biases in this system, the Wisconsin Supreme Court ruled in *State v. Loomis* that algorithms can indeed be used to sentence defendants and, by extension, that such sentences cannot be challenged on the basis of the use of such an algorithm because the algorithm is used only as part of the

Correctional Offender Management Profiles for Alternative Sanctions (COMPAS), *in* HANDBOOK OF RECIDIVISM RISK/NEEDS ASSESSMENT TOOLS 49 (Jay P. Singh et al. eds., 2018).

[45] *See* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES, Jan. 17, 2018, at 1.

[46] *See* Danielle Kehl, Priscilla Guo & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, RESPONSIVE COMMUNITIES INITIATIVE, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y, HARV. L. SCH., 2017, at 3, 9.

[47] Brennan & Dietrich, *supra* note 44, at 49, 52.

[48] *See* Larson et al., *supra* note 43.

[49] *See* Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, CORNELL U. LIBR. 2, 14 (Feb. 28, 2017), http://arxiv.org/abs/1703.00056 [https://perma.cc/7VAG-GY2Q].

[50] *Id.* at 5.

[51] *See generally The Northpoint Suite*, EQUIVANT, http://www.equivant.com/solutions/case-management-for-supervision [https://perma.cc/8MHX-26HN].

[52] *See* O'NEIL, *supra* note 11, at 209–10.

decision.[53] Based on the due process rule,[54] the defendant Loomis argued that the proprietary nature of the software prevented a challenge to its scientific accuracy and the data used.[55] He also asserted that the validity of the factors used to return risk scores could include possible impermissible sentencing factors, such as gender.[56] Loomis sought certiorari from the U.S. Supreme Court, which it denied in June 2017.[57] Thus, COMPAS still remains intact under the Wisconsin Supreme Court decision.

Lum and Isaac examined bias in a predictive policing system (*PredPol*) that was developed to flag areas where crimes may occur.[58] It appeared that the data fed into the *PredPol* algorithm were already biased: police arrests for drug crimes were disproportionately located in nonwhite areas, even though drug crimes were estimated to be distributed throughout the city in question.[59] Lum and Isaac then showed that, by training the predictive algorithm on these data, the algorithm inappropriately flags people from underrepresented groups as at risk of committing a crime.[60]

New York City uses *Palantir*, another system with which the same difficulties have been observed. The tool at issue allowed data from multiple sources to be analyzed and thereby predicted where

[53] *See* State v. Loomis, 881 N.W.2d 749, 753–54 (Wis. 2016).

[54] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1251, 1281 (2008).

[55] *See Loomis*, 881 N.W.2d at 757, 760; *see also* Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 99 (2014); Citron, *supra* note 54, at 1254, 1279.

[56] *Loomis*, 881 N.W.2d at 757. The court briefly raised concerns over how COMPAS' risk factor assessment may improperly correlate with the impermissible sentencing factor of race, before ultimately finding that COMPAS scores can still be used in sentencing. *See id.* at 763–764 ("Providing information to sentencing courts on the limitations and cautions attendant with the use of COMPAS risk assessments will enable courts to better assess the accuracy of the assessment and the appropriate weight to be given to the risk score."). Chief Justice Roggensack's concurrence in this case also cites race as an impermissible sentencing factor. *See id.* at 773 (Roggensack, J., concurring).

[57] Loomis v. Wisconsin, 137 S. Ct. 2290, 2290 (2017).

[58] *See* Kristian Lum & William Isaac, *To Predict and Serve?*, SIGNIFICANCE, Oct. 2016, at 14, 17.

[59] *See id.*; *see also* ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 65, 67, 69–75 (2017).

[60] *See* Lum & Isaac, *supra* note 58, at 19.

crimes were likely to occur.[61] A recent case, decided on December 27, 2017 by the Supreme Court of the State of New York, addressed a dispute between the Brennan Center for Justice at New York University School of Law and the New York City Police Department (NYPD) and involved a challenge to the algorithm's opacity.[62] Petitioner based its request on the Freedom of Information Law (FOIL) (Article 6 of the New York State Public Officers Law) and invoked the public's significant interest in the transparency of predictive policing systems.[63] The City of New York responded that the NYPD has to respect the vendor's trade secret and nondisclosure agreement.[64] Furthermore, disclosure of the predictive policing products' test results would discourage potential vendors from contracting with the NYPD and thereby limit the pool of technology available to it.[65] Nevertheless, the Supreme Court of the State of New York asked the NYPD to disclose the output data from the predictive policing system starting from six months before the date of the decision but rejected the request for disclosure of the input data.[66] This decision is a first step toward more transparency.

B. *Need for More "Ethical" Algorithms and Automated Decision-Making*

Algorithms and machine learning should not be viewed solely from an engineering perspective.[67] Such an approach must be complemented by a cognitive and human perspective with social considerations.[68] The lack of algorithms' oversight is socially

[61] *See* Peter Waldman, Lizette Chapman & Jordan Robertson, *Palantir Knows Everything About You*, BLOOMBERG (Apr. 19, 2018), https://www.bloomberg.com/features/2018-palantir-peter-thiel [https://perma.cc/KZ8Q-MMNA].

[62] *See* Brennan Ctr. for Justice at N.Y. Univ. v. N.Y.C. Police Dep't, 2017 N.Y. Misc. LEXIS 5138, at *5 (N.Y. Sup. Ct. Dec. 22, 2017).

[63] *See Brennan Ctr.*, 2017 N.Y. Misc. LEXIS 5138, at *5 (citing New York Public Officers Law § 87(2)).

[64] *See Brennan Ctr.*, 2017 N.Y. Misc. LEXIS 5138, at *8.

[65] *See Brennan Ctr.*, 2017 N.Y. Misc. LEXIS 5138, at *8–9.

[66] *See Brennan Ctr.*, 2017 N.Y. Misc. LEXIS 5138, at *21.

[67] Citron & Pasquale, *supra* note 4, at 6–7.

[68] *Id.*

unacceptable.[69] In this context, there is a social need for more fairness, accountability, and transparency of the algorithms[70] to challenge the biases and opacity of the results. Scholars, civil society organizations, and policymakers are increasingly asking for more algorithmic accountability, especially where individual decisions are solely based on an automatic system used by public agents. The need for "ethics of algorithms"[71] is observable in Europe[72] and also in the United States.[73] Some scholars associate algorithms with

[69] *Id.* at 8.

[70] Discussions are already underway in this arena—for example, the FAT conference (Fairness, Accountability, Transparency) on algorithmic systems is a multi-disciplinary conference that brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems. *See ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*)*, FAT CONFERENCE, https://fatconference.org/index.html [https://perma.cc/X62U-26FR].

[71] *See* Brent Daniel Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, BIG DATA & SOC'Y, July–Dec. 2016, at 1; Mireille Hildebrandt, *The New Imbroglio—Living with Machine Algorithms*, *in* THE ART OF ETHICS IN THE INFORMATION SOCIETY: MIND YOU 56, 57–58 (Liisa Janssens ed., 2016).

[72] *See, e.g.*, MIHALIS KRITIKOS, SCIENTIFIC FORESIGHT UNIT, EUROPEAN PARLIAMENTARY RESEARCH SERVICE, WHAT IF ALGORITHMS COULD ABIDE BY ETHICAL PRINCIPLES? 1 (2018) (providing as a European example the reinforcement of stigmatization of certain populations through measures taken by local councils in the UK which use algorithms to bring certain families to the attention of child protective services); *see also* Hildebrandt, *supra* note 26, at 41 (for instance, the draft GDPR included a provision on "the right to object and profiling," recognizing a right to object to automated decisions to protect against the possibility of being unethically profiled by algorithms).

[73] The City of New York enacted a local law on automated decision systems used by agencies on January 11, 2018 (returned unsigned by the Mayor on January 17, 2018). *See* N.Y.C. Local Law No. 49; *see also File #: Int 1696-2017*, N.Y.C. CITY COUNCIL, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 [https://perma.cc/4YWC-T538]. According to N.Y.C. Local Law No. 49, "'automated decision system' means computerized implementations of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions . . . concerning rules, policies or actions implemented that impact the public." N.Y.C. Local Law No. 49. The scope of this law is broad: it includes the use of algorithms, including artificial intelligence and machine learning processing. The purpose is to make or assist a decision. *Id.* The system has to be used by an agency appointed by the mayor in compliance with section 1-112 of the administrative code of the city of New York and the decision has to have an impact on the public. *Id.* This law doesn't yet furnish some provisions to regulate algorithms. It only states the creation of a task force, which was nominated on May 16, 2018 and will explore how New York City uses algorithms. *See* Press Release, Office of the Mayor, The City of New York, Mayor de Blasio Announces First-In-Nation Task Force to Examine Automated Decision Systems Used by the City (May 16, 2018), *available at*

six types of ethical problems: inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, negative transformative effects, and lack of traceability.[74] Broadly speaking, more transparency, fairness, and accountability are required.

Early on, the need for more transparency was demanded from the creators of algorithms. Nevertheless, one can easily understand that this is not relevant to governing algorithms because seeing does not mean knowing.[75] Seeing the inner workings of a system does not lead to understanding and controlling it. Plus, examining the code or pseudo-code would lead to a de-contextualization of the algorithm, which can frequently mutate. Although it is helpful to figure out how an existing technology works through reverse engineering, this process misses how the technology came to be this way (i.e., the socio-cultural embedding of code).[76] Besides, there are often technical limitations to a systematic approach because of the system's owners.

Consequently, our goal is not to consider the ways to "open the box." First, it may not be technically useful, because the algorithms are increasingly complex, especially the artificial intelligence systems. Moreover, the instructions could be unsupervised by programmers and hardly understandable for people. Second, having access to the algorithms once is not relevant if the instructions

https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by [https://perma.cc/JM3K-7YA4]. It is the first of its kind in the United States, and it will work to develop a process for reviewing "automated decision systems," commonly known as algorithms, through the lens of equity, fairness and accountability. *Id.* It will provide some recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address instances where people are harmed by agency automated decision systems. *See* N.Y.C. Local Law No. 49. More precisely, it aims to produce a report in December 2019 recommending criteria to determine which agency of the City is concerned and how implement procedures for reviewing and assessing City algorithmic tools to ensure equity and opportunity. *Id.*

[74] Mittelstadt et al., *supra* note 71, at 4–5.

[75] *See* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, NEW MEDIA & SOC'Y 973, 980 (2016).

[76] *See* Nick Seaver, *On Reverse Engineering: Looking for the Cultural Work of Engineers*, MEDIUM (Jan. 27, 2014), https://medium.com/anthropology-and-algorithms/on-reverse-engineering-d9f5bae87812 [https://perma.cc/D8MS-XZ6Z].

change without predictability. Third, because of trade secret and intellectual property laws, this request may not be legally permissible in many cases. Consequently, the social need for more knowledge and understanding requires consideration of the purposes of fairness and accountability.

Some legal scholars consider the notion of "fairness" an answer to requests for social justice.[77] Nevertheless, one can observe some diverging conceptions of just how algorithms achieve it.[78] The fairness of algorithms depends on their objectives.[79] Even if decisions are statistically derived and made consistently, actual fairness is not always achieved.[80] Moreover, even if an accurate algorithm exists, it "leads to generalizations about particular groups."[81] For instance, an algorithm "comes to the blanket conclusion that men tend to deserve higher risk scores than women."[82] "[W]ould it be fair [or even legal] for individuals to be judged based on immutable characteristics such as gender?"[83] Consequently, fairness by itself is not the best way to answer the need for less discriminatory algorithms.

Accountability starts with an agent and the outcome of its actions; the data holder (controller or processor) is accountable for ensuring compliance with the principles (and rights of the data subject).[84] The data holder is also supposed to have a mechanism in place to ensure compliance. Assumptions about computing and features of situations in which computers produce outcomes create four barriers to accountability: many people collaborate on systems

[77] *See* Kehl et al., *supra* note 46, at 30.

[78] *See id.*

[79] *See id.*

[80] *See id.*

[81] *Id.*

[82] *Id.*

[83] *Id.*

[84] *See* Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119), recitals 74, 79–80, at 14, 15–16 [hereinafter General Data Protection Regulation].

("problem of many hands"); errors tend to be pervasive and inevitable ("problem of bugs"); the temptation for "blaming the computer" is strong; and software ownership is not accompanied by liability.[85]

Even if the relevance of the principle of accountability is reduced in a computerized context, the need for accountability is more and more ripe.[86] Some actions decided by algorithms (i.e., automated decision-making) cause harms (or contribute significantly to causing them), and actions guided by faulty decisions or intentions (i.e., actions involving recklessness or negligence) should result in the data holder being held accountable or, eventually, liable.

I argue that current ethical requirements are too vague to enforce fair and compliant behavior of these automated decision-making tools' users. Self-regulation is not powerful enough to address these issues. Clear and binding rules are needed to fight against discrimination risks, on the one hand, and, on the other, to ensure the accountability of such automated decisions.

II. EU LEGAL FRAMEWORK ON AUTOMATED DECISION SYSTEMS

The European Union enacted a framework on automated decision-making (Article 22) in the General Data Protection Regulation 2016/679 (the "GDPR") on April 27, 2016.[87] In Section A, this Article discusses the GDPR's provisions for measures on

[85] *See* Helen Nissenbaum, *Accountability in a Computerized Society*, 2 SCI. & ENG'G ETHICS 25, 25 (1996).

[86] Paul B. de Laat, *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*, 31 PHIL. & TECH. 525, 526–27 (2018).

[87] The GDPR introduces some new provisions to address the risks inherent to profiling and automated decision-making. Statutory regulation of "automated individual decisions" by European data protection is not new. It was previously and explicitly addressed in Article 15 and Recital 41 of the Data Protection Directive 95/46/EU. *Compare* General Data Protection Regulation, *supra* note 84, recital 41, art. 15, at 8, 43, *with* Directive 1995/46/EC, of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data 95/46/EC, art. 15, 1995 O.J. (L 281) 43 [hereinafter Directive 95/46/EC]. Nevertheless, the GDPR extends the protection against decisions made solely by automated processing to cover profiling of data subjects, and more generally, any other form of automated processing. *See* General Data Protection Regulation, *supra* note 84, recital 41, art. 15, at 8, 43.

civil and commercial matters, while Section B highlights the Directive 2016/680/EU concerning criminal matters. This Part will conclude with a discussion of the Article 29 Working Party's Guidelines on automated decision-making and profiling, which completes the EU legal provisions on automated decision systems.

A. EU Legal Framework on Civil and Commercial Matters (GDPR)

This Section focuses first on the rights of data subjects that the GDPR strengthens. The second sub-section will outline some exceptions to such rights, with their attendant safeguards.

1. Rights of the Data Subject

a) Rights to Be Informed (Articles 13 and 14) and to Access (Article 15)

The data subject has several rights to be informed. They have (i) the right to know the existence of an automated decision-making system and that such system is used for his situation; (ii) the right to receive meaningful information concerning the logic involved; and (iii) the right to receive meaningful information on the significance and the contemplated consequences for his situation.[88] First, regardless of whether the data subject's personal data are collected from the data subject himself (Article 13, Section 2) or not (Article 14, Section 2), the controller shall provide the data subject the necessary information to ensure fair and transparent processing.[89]

Given the fact that the GDPR is founded on the core principle of transparency, controllers must ensure that they explain clearly and simply to individuals how the profiling or automated decision-making process works. In particular, where the processing involves profiling, the basis of such profiling must be made clear to the data subject. Furthermore, Article 15, Section 1 states that "[t]he data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or

[88] *Id.* at 41–42. Whether personal data related to a data subject are collected from the data subject (art. 13 § 2) or not (art. 14 § 2), the controller shall provide the data subject with some information necessary to ensure fair and transparent processing. *Id.*

[89] *Id.*

her are being processed, and, where that is the case, access to the personal data and . . . information."[90] What information the data subject has access to is of particular concern. The language of the GDPR indicates that the data subject should have access to "meaningful information about the logic involved, as well the significance and envisaged consequences of such processing for such data subject[s]," particularly in those cases where automated decision-making exists, such as those referred to in Article 22, Sections 1 and 4.[91] According to the Article 29 Working Party, "the controller has a duty to make available the data used as input to create the profile as well as access to information on the profile and details of [the] segments" of the data.[92] Nevertheless, Recital 63 provides some protection for controllers concerned about revealing trade secrets or intellectual property and, in particular, the copyright protecting the software, which may be particularly relevant in relation to profiling.[93] However, the Article 29 Working Party has reasoned that "controllers cannot rely on the protection of their trade secrets as an excuse to deny access or refuse to provide information to the data subject."[94]

b) Rights to Rectification and Erasure (Articles 16 and 17)

Profiling can involve an element of prediction, which increases the risk of inaccuracy. The input data may be inaccurate, or irrelevant, or taken out of context. There may be something wrong with the algorithm used to identify correlations. For example, Article 16 might apply where an individual is placed into a category that reveals something about his or her ability to perform a task, and such profiling is based on incorrect information. "Article 16 also provides

[90] *Id.* at 43.

[91] *Id.*

[92] ARTICLE 29 DATA PROTECTION WORKING PARTY, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679, WP 251, at 17, (Feb 6. 2018) [hereinafter Working Party Guidelines]. Article 29 Working Group is composed of National Data Protection Authorities under the Directive 95/46/EC, Article 29 (up to May 25, 2018). Since this date, this Group becomes the European Data Protection Board. *See All of the Article 29 Working Party Guidelines, Opinions, and Documents*, IAPP, https://iapp.org/resources/article/all-of-the-article-29-working-party-guidelines-opinions-and-documents/ [https://perma.cc/A8BD-AUFK].

[93] *See* General Data Protection Regulation, *supra* note 84, recital 63, at 12.

[94] Working Party Guidelines, *supra* note 92, at 17.

to the data subject the right to [supplement] the personal data with additional information."[95] Finally, the "rights to rectification and erasure apply to both the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person)."[96]

c) Right Not to be Subject to an Automated Decision (Article 22)

Article 22, Section 1 of the GDPR concerns "[a]utomated individual decision-making, including profiling."[97] In principle, the first paragraph states that "[t]he data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[98] This right is directly linked with the right to know (Articles 13-15): to exercise the right not to be subject to an automated decision, the data subject first needs to know if he is subject to it.[99] The right provided by Article 22 supposes three conditions: (i) a decision was made that is (ii) based solely on automated processing and that (iii) has legal effects or similarly significant consequences.[100] Examples of this are automatic refusal of an online credit application and e-recruiting practices without any human intervention (Recital 71).[101] "The controller cannot avoid the Article 22 provisions by fabricating human involvement."[102] For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing:

> To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the

[95] *Id.* at 18.

[96] *Id.*

[97] General Data Protection Regulation, *supra* note 84, art. 22, at 46.

[98] *Id.*

[99] *See id.* at 40–43, 46. Consequently, Articles 22 and Articles 13–15 must be understood all together.

[100] *See* Working Party Guidelines, *supra* note 92, at 20–22.

[101] *See* General Data Protection Regulation, *supra* note 84, at 14.

[102] *See* Working Party Guidelines, *supra* note 92, at 21.

> authority and competence to change the decision. As part of the analysis, [he or she] should consider all the relevant data.[103]

Despite the term "right," the Article 29 Working Party considers that Article 22 does not apply just when actively invoked by the data subject.[104] "Article 22 [establishes] a general prohibition" on individual decision-making, including profiling, "based solely on automated processing."[105] Consequently, individuals are automatically protected from the potential effects that this type of processing may have.

2. Broad Exceptions to the Rights

Article 22, Section 2 provides for three exceptions to the right not to be subject to an automated decision[106]: (a) if such decision is necessary under a contract;[107] (b) if such decision is authorized by European Union or Member State laws; and (c) if a data subject explicitly consents to the decision.[108] Otherwise, the EU rule provides for a default right not to be subject to automated decision-making. In the United States, the assumption is that a company or agency or person can use algorithmic decision-making however it wants, unless specifically prohibited by some rule.[109] For this reason, EU law seems to provide a better framework for protecting data subjects than U.S. law.[110]

[103] *Id.*

[104] *Id.* at 19.

[105] *Id.*

[106] General Data Protection Regulation, *supra* note 84, art. 22, at 46 (stating that "suitable measures to safeguard the data subject's rights . . . freedoms . . . and legitimate interests" have to be in place when the exceptions apply).

[107] *Id.* ("If the decision: (a) is necessary for the entering into, or performance of, a contract between the data subject and a data controller.").

[108] *Id.* The first two exceptions were previously provided by the Directive 95/46/EC (art. 15). *See* Directive 95/46/EC, *supra* note 87, at 43. Such exceptions are broad.

[109] *See, e.g.*, Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, § 104, 122 Stat. 881, 901 (2008). Here, the default is flipped: you cannot do it unless the nation has specifically permitted it. *See* General Data Protection Regulation, *supra* note 84, art. 22(2)(b), at 46.

[110] Defaults matter, empirically, and the burden to specifically allow something is much higher than the burden of doing nothing.

a) Performance of a Contract

"Controllers may wish to use solely automated decision-making processes for contractual purposes because . . . routine human involvement can sometimes be impractical or impossible due to the sheer quantity of data being processed."[111] "The controller must be able to show that this type of processing is necessary, taking into account whether a less privacy-intrusive method could be adopted."[112] Otherwise, "it would not be 'necessary'" and therefore not justified.[113]

b) Decision Authorized by Union or Member State Law

The automated decision-making has to be "expressly authorized by [a] Union or Member State law to which the controller is subject. . . ."[114] Such automated decision-making includes, for instance, "fraud and tax-evasion monitoring and prevention purposes."[115]

c) Explicit Consent

This new exception has to be defined according to Article 4, Section 11.[116] A specific consent supposes that the data subject understands the existence and meaning of automated decision-making and the envisaged consequences for his or her situation.

d) Safeguards to the Exceptions as Rights

In comparison to Article 15 of the Directive 95/46/EC, Article 22, Section 3 of the GDPR sets forth new guarantees.[117] When the

[111] Working Party Guidelines, *supra* note 92, at 23.

[112] *Id.*

[113] *Id.*

[114] General Data Protection Regulation, *supra* note 84, recital 71, at 14.

[115] *Id.* The prevention purposes have to be "conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller . . . ." *Id.*

[116] Section 11 states that the "'consent' of the data subject [is] any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her . . . ." *Id.* art. 4(11), at 34.

[117] *Compare* Directive 95/46/EC, *supra* note 87, at 43, *with* General Data Protection Regulation, *supra* note 84, at 46.

exceptions apply, "the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests."[118] These rights are a non-exhaustive list of "suitable measures."[119] The controller has to respect, at a minimum, the "right to obtain human intervention," the right for the data subject to "express his or her point of view," and the right "to contest the decision."[120] These requirements could be justified by one of the purposes of the GDPR—to improve the protection based on Article 8 of the EU Charter of Fundamental Rights.[121] Nevertheless, the right to contest the decision is not a right to reconsider it. Furthermore, human intervention is a key element, and any review must be carried out by someone who has the appropriate authority and capability to change the decision. Otherwise, this right would be useless. The reviewer should undertake a thorough assessment of all relevant data, including any additional information provided by the data subject.

B. EU Legal Framework on Criminal Matters

Council Directive 2016/680 was enacted the same day as the GDPR.[122] This text repealed the Council Framework Decision 2008/977/JHA.[123] For the first time in criminal law, Article 11 limits automated individual decision-making.[124] Paragraph 1 states that

[118] *Id.* art. 22, at 46.

[119] *Id.*

[120] *Id.*

[121] *Id.* recital 1, at 1.

[122] Council Directive 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of Such Data, and Repealing Council Framework Decision 2008/977/JHA, 2016 O.J. (L 119) 90 (EU) (elaborating on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection, or prosecution of criminal offenses or the execution of criminal penalties, and on the free movement of such data) [hereinafter Council Directive 2016/680].

[123] *Id.* at 89.

[124] *See id.* at 109. Before the Directive, a decision-cadre was enacted but did not contain any provision concerning the automated decision-making. *See* Council Framework Decision 2008/977/JHA of 27 November 2008 on the Protection of Personal Data Processed in the Framework of Police and Judicial Cooperation in Criminal Matters.

"Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her, to be prohibited . . . ."[125] As opposed to Article 22, Section 1 of the GDPR, this provision is a prohibition governing the data controller and not a right afforded to the data subject, which is quite different.[126] Specifically, there are fewer guarantees in the case of a breach of the law.

Article 11, Section 1 also provides some exceptions if the automated individual decision-making is "authori[z]ed by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller."[127] This exception and its safeguards are the same as in the GDPR Article 22, Section 2(b).[128] Paragraph 2 does not authorize decisions that should be based on sensitive data, except if suitable measures are in place to safeguard the data subject's rights, freedoms, and legitimate interests.[129] Nevertheless, the profiling of natural persons based on sensitive data through which they can be discriminated against is prohibited.[130]

The GDPR and the Directive provide a legal framework to limit automated decision-making but are completed by Guidelines from the Article 29 Working Party.

[125] Council Directive 2016/680, *supra* note 122, at 109.

[126] *Compare* General Data Protection Regulation, *supra* note 84, at 46, *with* Council Directive 2016/680, *supra* note 122, at 109.

[127] Council Directive 2016/680, *supra* note 122, at 109.

[128] *Compare* General Data Protection Regulation, *supra* note 84, at 46, *with* Council Directive 2016/680, *supra* note 122, at 109.

[129] Council Directive 2016/680, *supra* note 122, at 110.

[130] *Id.* Article 11, section 3 is in accordance with articles 21 and 52 of the EU Charter of Fundamental Rights. Recital 38 adds that the safeguards should include the provision of specific information to the data subject and the right to obtain human intervention, in particular to express his or her point of view, to obtain an explanation of the decision reached after such assessment or to challenge the decision. *Id.* at 95.

C. Guidelines of the Article 29 Working Party on Automated Decision-Making and Profiling

The Article 29 Data Protection Working Party, adopted on February 6, 2018, sets forth Guidelines on automated decision-making and profiling for the purpose of Regulation 2016/679.[131] Among all the recommendations, I will focus on the transparency and fairness requirements.

The GDPR only defines "profiling," which is related to automated decision-making. According to Article 4, Section 4, profiling is:

> [A]ny form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements.[132]

Consequently, profiling is composed of three elements: (1) an *automated* form of processing (2) carried out on *personal data*, (3) the

[131] Working Party Guidelines, *supra* note 92, at 2. The Guidelines reveal the issues concerning the fairness, transparency, and accountability requirements for algorithms. *Id.* at 6. The Guidelines note:

> The GDPR introduces provisions to ensure that profiling and automated individual decision-making (whether or not this includes profiling) are not used in ways that have an unjustified impact on individuals' rights; for example:
>
> - specific transparency and fairness requirements;
> - greater accountability obligations;
> - specified legal bases for the processing;
> - rights of individuals to oppose profiling and, specifically, profiling for marketing; and,
> - if certain conditions are met, the need to carry out a data protection impact assessment [DPIA].

Id. at 6.

[132] General Data Protection Regulation, *supra* note 84, art. 4, at 33.

object of which is to *evaluate personal aspects* about a natural person.[133] Article 4, Section 4 refers to "any form of automated processing" rather than "solely automated processing" (referred to in Article 22).[134]

The GDPR states that "profiling" is the "automated processing of personal data [for] evaluating personal aspects" and, in particular, for analyzing or making predictions about individuals.[135] The use of the word "evaluating" suggests that profiling involves some form of assessment or judgments about a person. According to the Article 29 Working Party guidelines, profiling means "gathering information about an individual (or a group of individuals) and evaluating their characteristics or behavioral patterns" in order to categorize them, and to analyze and/or make predictions about their ability to perform a task, their interests, or their likely behavior.[136] For instance, the data broker compiles the data collected from different public and private sources to develop profiles on the individuals and places them into segments that outline important aspects of consumer needs, consumer behavior, brand preferences, product usage levels, and so on. The data broker sells this information to companies who wish to improve the targeting of their goods and services. He carries out profiling by placing a person into a certain category according to his or her interests.

Whether something is "automated decision-making," as defined in Article 22, Section 1 will depend upon the circumstances. Indeed, "automated decision-making" has a different scope than profiling and its results may partially overlap with, or result from, profiling. "Solely automated decision-making is the ability to make decisions by technological means without human involvement."[137] Automated decisions can be made with or without profiling, and profiling can take place without making automated decisions. However, profiling and automated decision-making are not necessarily separate activities. Something that starts off as a simple automated decision-

[133] *See* Working Party Guidelines, *supra* note 92, at 6–7.

[134] *Id.* at 7.

[135] General Data Protection Regulation, *supra* note 84, recital 71, at 14.

[136] Working Party Guidelines, *supra* note 92, at 7.

[137] *Id.* at 8.

making process could become a process based on profiling, depending upon how the data is used. Decisions that are not solely automated might also include profiling. For example, before granting a mortgage, a bank may consider the credit score of the borrower, with humans carrying out additional meaningful intervention before any decision is applied to an individual.

Finally, according to these guidelines, there are three potential ways in which profiling may be used: "(i) general profiling; (ii) decision-making based on profiling; and (iii) solely automated decision-making," (including profiling) which may legally affect the data subject, or otherwise significantly affects the data subject.[138] Additional safeguards and restrictions apply in this third case.

The GDPR and the Directive provide a legal framework to address the social need for accountability of automated decision-making. Generally speaking, these rules are protective of data subjects. Nevertheless, I observe several limits.

III. Limits of the EU Legal Framework on Automated Decision-Making

Despite the goal of protecting the data subjects and promoting the understanding of the issues by EU lawmakers, I argue that the given solutions are insufficient to improve the previous rules and protect the vulnerable populations against the risks of opacity and discrimination of algorithms. Specifically, the exceptions afford too much flexibility in favor of the private and public players as well as the Member States, based on the GDPR and Directive 2016/680/EU. Consequently, this protection is too weak and too diverse. Each of these shortcomings will be addressed in turn.

A. A Weak Protection Related to Automated Decision-Making and Profiling

The provisions contain many internal limits to the protection of data subjects. However, personal data legislation is not the only way to achieve the goal of protecting natural persons against algorithmic

[138] *Id.*

discrimination. Other fields of law have to be considered. Consequently, some external limits must be taken into account.

1. Internal Limits of the EU Personal Data Legislation

a) Limits Concerning the Right to Have Meaningful Information About the Logic Involved

The first difficulty is understanding how to satisfy the requirement of having "meaningful information about the logic involved,"[139] especially in cases where a machine learning process involves multiple data sources, dynamic development, and elements that are opaque, whether for technological or proprietary reasons.[140] The growth and complexity of machine learning can make it challenging to understand how an automated decision-making process or profiling works. One should evaluate what will constitute "meaningful information" about "logic" from the perspective of the data subject. As shown above, disclosure of the algorithms' full code and detailed technical descriptions of machine learning processes are unlikely to help. "A high-level, non-technical description of the decision-making process is more likely to be meaningful."[141] Moreover, intellectual property ("IP") rights and trade secrets create some barriers, and neither the GDPR nor the Directive provide exceptions or limitations to the scope of such proprietary rights.[142] A potential conflict of legal norms between IP rights and data protection rights resolves in favor of the former.

According to the Article 29 Working Party, "[t]he controller should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision."[143] The GDPR requires the controller to provide meaningful information about the logic involved, but not necessarily a complex explanation

[139] General Data Protection Regulation, *supra* note 84, art. 15(1)(h), at 43.

[140] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1373–74 (2018).

[141] Kuner et al., *supra* note 40, at 2.

[142] The GDPR and the Directive state only the respect for intellectual property rights without provision to conciliate them with the requirement of transparency. *See, e.g.*, General Data Protection Regulation, *supra* note 84, recital 63, at 12.

[143] Working Party Guidelines, *supra* note 92, at 25.

of the algorithms used or disclosure of the full algorithm.[144] The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision.[145] Nevertheless, such provisions cannot give guarantees against biases and discrimination, and the data subject only knows the consequences of such systems and ignores potential biases. Moreover, one has no way to prove the existence of biases or avoid them. Finally, these measures are insufficient to avoid the main risks of algorithms concerning biases and discrimination.

Besides, there is a debate among scholars as to whether Articles 13–15 and 22 of the GDPR provide the right to an explanation.[146] A reconciliation[147] or explanation[148] can be found, but the explanation is not necessary.[149] Indeed, meaningful information about the logic involved does not mean a right to an explanation. It does not provide the data subject with an individual right to know and understand what exactly happened to him. Nevertheless, the "suitable measures" of Article 22, Section 3 are not an exhaustive list of rights, and "[t]he only other right that might benefit a data subject would be a right to be given an explanation for an automated decision."[150] The explicit mention of this right in the GDPR occurs

[144] *See* General Data Protection Regulation, *supra* note 84, recital 58, at 11 (stating that the principle of transparency is "of particular relevance in situations where the proliferation of actors and the technological complexity of practice makes it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising.").

[145] *See id.*

[146] *See, e.g.*, Andrew D. Selbst & Julia Powles, "*Meaningful Information*" *and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233, 234 (2017). *Compare* Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation,"* AI MAG., Fall 2017, at 50, 55–56, *with* Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 77 (2017).

[147] *See* Selbst & Powles, *supra* note 146, at 241–42.

[148] *See, e.g.*, Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 192–93 (2019).

[149] *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a* "*Right to Explanation*" *Is Probably Not the Remedy You are Looking for*, 16 DUKE L. & TECH. REV. 18, 21, 81 (2017).

[150] Isak Mendoza & Lee A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 8, 15 (Tatiani

only in Recital 71, which is not binding.[151] It states that: "[S]uch processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain . . . an explanation of the decision reached after such assessment and to challenge the decision."[152] Consequently, there may be a tension between the Article 22 right to obtain general information about a decision-making process and the right "to obtain . . . an explanation of the decision reached after such assessment and to challenge the decision," included in Recital 71.[153] According to some scholars, "[a]lthough not directly binding, [Recital 71] may embolden regulators and courts to try to compel data controllers to provide explanations of specific outcomes in particular cases, and not merely 'meaningful information' about 'logic' in general."[154] However, because Recital 71 is not binding, it cannot be used as a basis to claim a right.[155] Moreover, nobody knows if the European Court of Justice will broadly interpret Article 22 pursuant to Recital 71. At this step, it seems too early to affirm the existence of such right to an explanation, even implicitly. Finally, this question is not the most relevant concerning the impact of the GDPR.[156]

b) Limits Concerning the Safeguards

The rights outlined in the GDPR do not include a right to an explanation. These rights merely afford the right to ask for a human being, and not a machine, with whom to interact, without ensuring a better understanding. Moreover, even if there is human intervention, it may not be feasible to conduct a meaningful review of a process. For instance, if the process may have involved third-party data and algorithms, pre-learned models, or inherently opaque

Synodiou et al. eds., 2017). *See also* General Data Protection Regulation, *supra* note 84, at 46.

[151] *See* General Data Protection Regulation, *supra* note 84, recital 71, at 14.

[152] *Id.*

[153] *Id.* at 14, 46.

[154] Kuner et al., *supra* note 40, at 2.

[155] *See* Mendoza & Bygrave, *supra* note 150, at 85.

[156] Roland Vogl, Brian Casey & Ashkon Farhangi, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 142, 151 (2019).

machine learning techniques, it may not be possible to inform data subjects about more than the machine.[157]

For its part, the right to contest a decision could be a sort of right to appeal. Nevertheless, the legal framework lacks any guarantees regarding a potential right to reconsider a decision, or even sufficient information to that effect. What will concretely happen if a data subject contests? Maybe he could lose the right to obtain a decision if he does not accept an automated one. In the event the subject gets a real second chance at obtaining another decision, there is no way to know whether the decision would be manual or automated, since no requirement as to the type exists. Affording the data subject the right to demand a manual re-examination of the decision would offer a higher level of protection.

Some limits to the right not to be subject to a decision based on automated processing (Article 22, Section 1 and Directive 2016/680/UE, Article 11) are observable in these provisions' terms. First, this right concerns the decision based "solely" on automated processing. This means that automated processing could be used without any restrictions, limitations, or guarantees if it is not the only means for making the decision. However, it is very easy to pretend that other processes are used to make a decision, although it would not be true. The lack of control prevents understanding of the decision. Moreover, decisions made by machines have a strong impact on human decisions. It is very difficult to make a different decision than the one suggested by the machine and to justify it. Second, a decision based on automated processing has to produce legal effects concerning the data subject, or otherwise to "significantly" affect them. However, the automated individual decision may have a negative or discriminatory impact without producing legal effects or significantly affecting the data subject. Moreover, what does "significantly" mean? It is difficult to draw the line between what is "significant" and not. It seems to require a high level of impact, although an impact that is not "significant" could have a very negative effect on the data subject.

[157] *See id.* at 185–86.

The "explicit consent" provision is an exception to the right not to be subject to a decision based solely on automated processing.[158] How should this exception apply? As Cate and others have said, "[t]o be sufficiently 'specific,' will a separate consent be required for each situation in which personal data are to be processed for automated decision-making, for example, in particular employment, financial, or medical contexts?"[159] Such interpretation may be too stringent for the data controller and not necessarily helpful to the data subject. Moreover, the overload of information can kill the meaning by obtaining explicit consent without being informed and being freely given. Indeed, "[i]t is standard practice, at least at the internet context, for companies to prompt data subjects to consent to various data-processing operations."[160] Besides, must the data controller provide an opportunity to revoke the consent? Moreover, even if we consider "an algorithmic process," which "can in theory be explained," how can we do that in a meaningful and intelligible way to a data subject to obtain a real consent?[161] Consequently, will it be meaningful for him? Finally, the prohibition concerning the sensitive data provided by Article 22, Section 4 also can be derogated by obtaining explicit consent. The inclusion of the exception for explicit consent impacts the data subject's interest.

To compare this with Article 15 of Directive 95/46/EC, Article 22 of the GDPR specifically accounts for profiling,[162] which was the subject of many debates during the adoption of this regulation.[163] Recital 71 provides some guarantees in case of error or discrimination.[164] These measures go in the right direction. Nevertheless, they

[158] General Data Protection Regulation, *supra* note 84, arts. 9(2)(a), 22, at 38, 46.

[159] Kuner et al., *supra* note 40, at 2 (Fred H. Cate, an Editor at *International Data Privacy Law*, is a co-author of this article).

[160] Mendoza & Bygrave, *supra* note 150, at 96.

[161] Kuner et al., *supra* note 40, at 1–2.

[162] *See* General Data Protection Regulation, *supra* note 84, art. 22, at 46 (providing some guidelines on automated decision-making and profiling for the purpose of Regulation 2016/679, which reveals the issues concerning the fairness, transparency and accountability requirements of algorithms). *Cf.* Directive 95/46/EC, *supra* note 87, at 43.

[163] *See* General Data Protection Regulation, *supra* note 84, at 14 (stating that profiling is "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person").

[164] *See id.* (stating "the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate

are only provided by an unbinding recital and not by the text itself. The same limit can be found at the end of Recital 71, which adds that such a measure on the profiling should not apply to children.[165] This wording is not reflected in the article itself, so this provision does not represent an absolute prohibition, as safeguards have to be in place and appropriate for children. Such provisions are essential. However, they may only have a potential influence on the future decisions of the European Court of Justice if the Court decides to use them.

Despite the GDPR's purpose of improving the protection of Europeans in a digital context and the new basis of the EU Charter of Fundamental Rights (Article 8), the protection does not seem so efficient. Article 22's ability to have a practical impact on automated profiling, particularly when applied to decisional systems that are complex and opaque, is also doubtful. Many activities and business models of the digital economy are based on massive data processing and algorithmic systems. Consequently, being compliant with the GDPR usually requires many changes in personal data processing in order to respect the rights of the data subjects.

to ensure, in particular, that factors which result in inaccuracies in personal data are corrected . . . and the risk of errors is minimised."). The controller also should:

> secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.

Id. These provisions focus on the sensitive data, and the recital adds that "automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions." *Id.*

[165] *See id.* at 14. Recital 38 states:

> Children merit specific protection with regard to their personal data, as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data. Such specific protection should, in particular, apply to the use of personal data of children for the purposes of marketing or creating personality or user profiles and the collection of personal data with regard to children when using services offered directly to a child.

Id. at 7.

2. External Limits of the EU Personal Data Legislation

First, beyond the content of a specific text, such as the GDPR, a personal data statute is not necessarily the best way to ensure protection against the biases and opacity of algorithms. Indeed, the material scope of this kind of legislation is traditionally the protection of "personal data" or "personally identifiable information" from natural persons, and the goal is to achieve this protection by giving them some individual rights.[166] For instance, Article 22 of the GDPR focuses on automated individual decision-making.[167] What about automated collective decision-making concerning a group? Discrimination toward a group of persons can also be observed. Moreover, an individual discriminatory decision is often taken with consideration of multiple criteria, such as a category of people (e.g., black people, young people, women). Consequently, the problem of discrimination is a global one that concerns not only a specific person but, more generally, some groups of people who represent the vulnerable populations. Personal data legislation cannot properly address the issue of algorithmic transparency, and this subject matter has to be considered separately.

Finally, there are challenges with the efficiency of personal data legislation in its interrelations with other fields of law. The need for more algorithmic transparency also has to be considered in light of competition law, consumer law, and, eventually, the constitutional law of other countries (e.g., First Amendment and free speech). All of these fields overlap to address the algorithmic problems. For more efficiency, the question has to be thought of in global terms. The same conclusion applies concerning the regulator's choices. It seems to be insufficient to give the data protection authorities the task of controlling the algorithms. In the EU, many of the Member States do not have the resources to do it seriously, especially with respect to providing an oversight for the more complex and opaque algorithms.

Finally, the efficiency of the GDPR will depend on the capacity of the EU Member States to create some processes and tools to

[166] *See* General Data Protection Regulation, *supra* note 84, art. 23, at 46.

[167] *See id.*

enforce the law. This sort of challenge afflicts not just Article 22 rights, but more broadly the provisions of the GDPR as well.

B. A Diverse Protection Related to Automated Decision-Making and Profiling

After the integration of the "personal data package" at the national level, I point out that the commonality of the frameworks between the Member States is smaller than expected. The GDPR affords many flexibilities to the Member States to determine their requirements (opening clauses) at the national level, such as Article 22, Section 2(b).[168] As some scholars have said, "Article 22, Section 2(b) opens up [the possibility] for a great deal of nationally authorised automated decisional processes with potentially differing standards [to] be[] applied from country to country, thereby undermining the harmonisation aims of the Regulation."[169] Without pretending to consider all of these national laws, I will study the implementation of the EU provisions on: (1) civil and commercial matters, and (2) criminal matters in several Member States (France, Germany, Ireland, and the UK). All of these legal frameworks provide different rules. Such diversity challenges the purposes of the GDPR and EU politics to build a "digital single market."

1. National Legal Frameworks on Civil and Commercial Matters

I am specifically studying the French law because the automated decision-making requirements were originally adopted in this country.[170] Moreover, the French government has announced its goal to improve the GDPR's level of protection in the data subject's favor, for instance, in terms of the right to an explanation on automated decision-making.[171] I am also briefly studying German, Irish, and English laws.

[168] General Data Protection Regulation, *supra* note 84, at 46.

[169] Mendoza & Bygrave, *supra* note 150, at 95.

[170] *See* Gianclaudio Malgieri, *Automated Decision-Making in the Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations*, COMPUTER L. & SECURITY REV., Oct. 2019, at 14.

[171] Loi 2018-493 du 20 juin 2018 relative à la protection des données personnelles, [Law 2018-493 of June 20, 2018 on the Protection of Personal Data], JOURNAL OFFICIEL DE LA

a) French Law: A Higher Level of Protection?

The "Digital Republic Act" (*Loi n° 2016-1321 pour une République numérique*) of October 7, 2016 anticipated some provisions of the GDPR (the "Data Protection Act" or "Act").[172] The new Data Protection Act was enacted on June 20, 2018.[173]

The Data Protection Act makes extensive use of the opening clauses to increase the level of the data subject's protection and modify Article 11 of the previous Law 78-17.[174] The Act also prohibits decisions solely based on automated-decision making.[175] Both principles provide a higher level of protection than the GDPR, which does not prohibit decisions made solely based on automated processing to predict or evaluate some of the data subject's personal details. It only provides a right not to be subject to such decisions (Article 22, Section 1).[176] Consequently, such French provisions are

RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], June 20, 2018, p. 1 (Fr.) [hereinafter Law 2018-493 of June 20, 2018 on the Protection of Personal Data].

[172] *See* General Data Protection Regulation, *supra* note 84, arts. 15–22, at 43–46 (discussing the right to be forgotten, the right to portability, the right for individuals to give instructions relating to the storage of data, and the right to erasure and disclosure of their personal data after their death). This Law increased the sanctioning powers of the Data Protection Authority (CNIL): the maximum fines were increased from €150,000 to €3 million in case of data protection infringements. *Id.*

[173] Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, at 9 (discussing how this law enables the implementation of the GDPR by updating the Data Protection Act of Jan. 6, 1978). The Data Protection Act of Jan. 6, 1978 is also known as "informatique et libertés." Loi 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [Law 78-17 of Jan. 6, 1978 on Information Technology, Data Files and Civil Liberties], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Jan. 6, 1978 (Fr.) [hereinafter Law 78-17 of Jan. 6, 1978 on Information Technology, Data Files and Civil Liberties]. The new law was challenged in front of the *Conseil constitutionnel/Constitutional Council*. *See* Conseil constitutionnel [CC] [Constitutional Court] decision No. 2018-765DC, June 12, 2018, Rec. 71 (Fr.).

[174] Moreover, the provisions on automated decision-making adopted by Directive 95/46/EC, repealed by the GDPR, were inspired by the Data Protection Act of Jan. 6, 1978. *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, at 14; *see also* Law 78-17 of Jan. 6, 1978 on Information Technology, Data Files and Civil Liberties, *supra* note 173. *Cf.* Directive 95/46/EC, *supra* note 87, at 43.

[175] *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, art. 21, at 14 (stating that "[n]o other decision having a legal effect on an individual" or similarly and significantly affecting him or her can be "taken solely based on [an automated] processing of data," including profiling).

[176] *See* General Data Protection Regulation, *supra* note 84, at 46.

more protective. Nevertheless, the Article 29 Working Party Guidelines interpret Article 22, Section 1 in the same manner.[177]

Some exceptions limit protections. The third paragraph of the Act follows Article 22, Section 2 of the GDPR, authorizing decisions taken solely based on automated processing when they fall under two mandatory exceptions—i.e., points (a) and (c): contracts and explicit consent, respectively.[178] This clause of the Act, by excluding any mention of exception (b) of Article 22, Section 2 of the GDPR, appears to reserve to other Member States the flexibility to enact other exceptions to their national laws.[179] The Data Protection Act further reiterates the GDPR's exceptions and the safeguards of the data subject's rights, freedoms, and legitimate interests (set forth in Article 22, Section 3 of the GDPR), providing for at least the right to obtain human intervention, the right for the data subject to express one point of view, and the right to contest the decision.[180] Moreover, the French law requires that "the rules defining this processing as well as the main characteristics of its implementation [be] provided to the data subject at his request, except the secrets protected by the law."[181]

The French legislature also used the opening clause, pursuant to Article 22, Section 2(b) of the GDPR, to create a new exception for administrative decisions.[182] In this case, the data subject has to be

[177] *See* Working Party Guidelines, *supra* note 92, at 20.

[178] *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, art. 21, at 14.

[179] *See id.*; *see also* General Data Protection Regulation, *supra* note 84, at 46.

[180] *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, at 14, 16, 20–21.

[181] *Id.* at 14.

[182] *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, art. 21, at 14. Surprisingly, this exception concerns the government, as mistrust of the vast governmental databases led to the original adoption of the French Data Protection Law (SAFARI project) in 1978, and there is probably no more trust in the government today. *See* Peter Sayer, *French Plan for Biometric Database of 60 Million People Sparks Outcry*, PCWORLD (Nov. 8, 2016, 5:51 AM), https://www.pcworld.com/article/3139461/french-plan-for-biometric-database-of-60-million-people-sparks-outcry.html [https://perma.cc/94BV-55DS]. This exception excludes the processing of sensitive data. *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, art. 21, at 14 (citing *id.*, art. 8, at 7).

informed about the use of an automated system[183] and has an explicit right to an individual explanation. According to the *Conseil constitutionnel* (the Constitutional Council), the data controller has to control the "algorithmic processing [as well as its developments] to explain, in details and in an intelligible [form]," to the data subject the way the processing was applied to his situation.[184] Such characteristics point out the machine learning methods relevant to the data subject.

In furtherance of the goal of achieving "accessibility and comprehensibility" of the law for data subjects,[185] trade secrets and IP rights have also been subjected to higher standards. The *Conseil constitutionnel* decided that, when the principles of the inner functioning of an algorithm cannot be communicated without infringing a secret or IP interest, no individual decision can be taken on the exclusive basis of this algorithm.[186] Such a rule has a significant impact because it is a way to reconcile, on the one hand, secrecy and property and, on the other hand, transparency and accountability.

These conditions stated by the *Conseil constitutionnel* reveal the need to consider the impact of tools and to check whether they are able to satisfy a legal requirement for transparency. Consequently, while algorithms that change their rules (e.g., machine learning and deep learning systems) have to be excluded, the algorithms that are protected by secrets or IP rights do not have to be excluded.

Besides, there is another problem that is not considered by the *Conseil constitutionnel*. One can also wonder how to monitor access to the rules defining the automated processing provided by consumer and personal data laws. Two different agencies have

[183] CODE ADMINISTRATIF [C. ADM.] [ADMINISTRATIVE CODE] art. L311-3-1 (Fr.) (It states that an individual decision taken on the basis of an algorithmic treatment includes an explicit mention by informing the data subject (i.e., the right to be informed)).

[184] Conseil Constitutionnel [CC] [Constitutional Court] decision No. 2018-765DC, June 12, 2018, J.O. 141, recital 71, at 13 (Adding that in cases of judicial remedy, the judge can ask the administration to explain to the data subject how the algorithm has been implemented. Moreover, in explaining to the subject the way processing was applied to his situation, the data controller cannot use, as an exclusive means for an individual administrative decision, algorithms able to change their own rules by themselves without the data controller's control and validation.).

[185] *Id.* recital 66, at 12.

[186] *Id.* recital 70, at 13.

jurisdiction: *Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes* ("DGCCRF") and the *Commission Nationale Informatique et Libertés* ("CNIL"), which is the French Data Protection Authority.[187] Neither has sufficient powers and human resources to ensure efficiency and algorithmic oversight.

b) Other European National Laws

i. Germany

The German lawmakers have made extensive use of Article 22, Section 2(b) of the GDPR in a different way than the French legislators. Germany renewed its Data Protection Law on June 30, 2017.[188] Section 37 concerns automated individual decision-making, including profiling, and first reiterates the safeguards of the GDPR[189] before stating in Paragraph 1 a single additional exception to the right not to be subject to a decision based solely on automated processing.[190] Under this new exception, decisions may be based on the processing of health data (Article 4, Section 15 of the GDPR).[191] This means that such exception applies in favor of the healthcare sector. It is too early to say what impact such a measure will have. Nevertheless, one can already observe that the choices made by the German and French lawmakers are wholly different.

[187] *See Consumer Rights in France*, ANGLOINFO, https://www.angloinfo.com/how-to/france/lifestyle/shopping/consumer-rights [https://perma.cc/XR5Q-QEHA].

[188] Bundesdatenschutzgesetz [BDSG] [Federal Data Protection Act], June 30, 2017, BUNDESGESETZBLATT [BGBL I] at 2097 (Ger.), *translation at* https://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.pdf [https://perma.cc/E8Q6-X7HE].

[189] *Id.* at ch. 2, sec. 37 ("In addition to the exceptions given in Article 22 (2) (a) and (c) of [the GDPR] . . . .)

[190] *Id.* at ch. 2, sec. 37 (The exception applies "if the decision is made in the context of providing services pursuant to an insurance contract and (1) the request of the data subject was fulfilled . . . or (2) the decision is based on the application of binding rules of remuneration for therapeutic treatment . . . ."). Section 37 also outlines remedies available to data subjects in the event their request is not granted in full: "[T]he [data] controller [shall take] suitable measures . . . to safeguard the data subject's legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision." *Id.* Finally, "the controller shall inform the data subject of these rights no later than the notification indicating that the data subject's request will not be granted in full." *Id.*

[191] *See id.*

ii. Ireland

Ireland enacted the Data Protection Act in 2018.[192] Article 52 concerns rights in relation to automated decision-making.[193] It also states some exceptions.[194] The exceptions are broad, and a lot of automated decisions based solely on automated processing could be authorized or required by or under an enactment in many circumstances. Moreover, even if these conditions are not required, the controller could use an automated decision if he adopts some measures to safeguard the data subject's legitimate interests. The law thus gives significant opportunity to the data controller. Furthermore, such safeguards must include the making of arrangements to enable one "to make representations to the controller in relation to the decision."[195] These exceptions seem neither clear nor stringent.

iii. United Kingdom

Despite Brexit, the United Kingdom enacted a Data Protection Act on May 23, 2018.[196] Chapter 2, Section 14 concerns the safeguards of automated decision-making authorized by law.[197] These

[192] *See* Data Protection Bill 2018 (Act. No. 7/2018) (Ir.), http://www.irishstatutebook.ie/eli/2018/act/7/enacted/en/pdf [https://perma.cc/44W4-G9FL].

[193] *Id.*

[194] *See id.* These exceptions are:

> where—
>
> (a) the decision is authorised or required by or under an enactment, and
>
> (b) either—
>
> > (i) the effect of that decision is to grant a request of the data subject, or
> >
> > (ii) in all other cases (*where subparagraph (i)* is not applicable), adequate steps have been taken by the controller to safeguard the legitimate interests of the data subject which steps shall include the making of arrangements to enable him or her to make representations to the controller in relation to the decision.

Id.

[195] *Id.*

[196] *See* Data Protection Act 2018, c. 12 (UK), http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [https://perma.cc/2U2N-YMMC].

[197] Chapter 2, section 14 (3) states:

> Where a controller takes a qualifying significant decision in relation to a data subject based solely on automated processing—
>
> (a) the controller must, as soon as reasonably practicable, notify the data subject in writing that a decision has been taken based solely on automated processing, and

measures provide a right of information, a right to reconsider, and a right to obtain another decision, which has to be manual or based on automated processing and which assists a decision-making.

The right to reconsider is an interesting right because it creates an opportunity to obtain another decision that could be more positive.[198] However, this second chance does not avoid the bias and opacity problems of algorithmic decision-making. Once again, the UK's solution is different from the French and German ones.

2. National Legal Frameworks on Criminal Matters

As opposed to a Regulation, a Directive has to be implemented in the national laws by the Member States. I will discuss the implementation of the Directive 2016/680/EU concerning criminal law protections of personal data in both the French Data Protection Act and in other countries.

a) French Law

Article 30, Section 1, Art. 70–9 of the French Data Protection Act addresses automated decision-making on criminal matters.[199] Paragraph 1 states a principle of prohibition concerning judicial

> (b) the data subject may, before the end of the period of 21 days beginning with receipt of the notification, request the controller to—
> (i) reconsider the decision, or
> (ii) take a new decision that is not based solely on automated processing.

Id.

[198] *Id.* at cl. 50.

> [I]f a request is made to a controller under subsection (2), the controller must, before the end of the period of 1 month beginning with receipt of the request—
> (a) consider the request, including any information provided by the data subject that is relevant to it,
> (b) comply with the request, and
> (c) by notice in writing inform the data subject of—
> (i) the steps taken to comply with the request, and
> (ii) the outcome of complying with the request.

Id.

[199] *See* Law 2018-493 of June 20, 2018 on the Protection of Personal Data, *supra* note 171, at 18.

decisions.[200] Paragraph 2 addresses other decisions.[201] Finally, paragraph 3 prohibits discrimination based on profiling.[202] This last paragraph is the same as the Directive.

The French provisions seem to prohibit predictive justice and predictive policing systems based exclusively on algorithmic decision-making. Such systems can only be used to assist decision-making. This interpretation is strict in a criminal matter, and the lawmaker's purpose is to protect the data subjects by refusing the use of this kind of tool. This solution is one of the more stringent ones enacted by an EU Member State. Even if decision-making is not solely based on algorithms, such tools may nevertheless substantially influence the decision-maker.

b) Other European National Laws

i. Germany

Section 54, paragraph 1 of the new Federal Data Protection Act on automated individual decision-making states an authorization principle.[203] Distinct from French law, this permits the use of such tools in a criminal matter, with exceptions.[204] Moreover, discrimination based on profiling is also prohibited.[205] Similar to a majority of European Member States, the German lawmakers use the same words as the Directive and authorize by law a decision based solely

[200] *Id.* ("No judicial decision involving an assessment of a person's conduct may be based on automated processing of personal data intended to assess certain aspects of that person's personality.").

[201] *Id.* ("No other decision which produces legal effects in respect of a person or significantly affects him may be taken solely on the basis of automated data processing intended to foresee or evaluate certain personal aspects relating to the person concerned.").

[202] *Id.* ("Any profiling which discriminates against natural persons on the basis of the special categories of personal data referred to in Article 8 (1) shall be prohibited.").

[203] Bundesdatenschutzgesetz [BDSG] [Federal Data Protection Act], 2097, at 1 (Ger.) ("A decision based solely on automated processing which produces an adverse legal effect concerning the data subject or significantly affects him or her shall be permitted only when authorized by law.").

[204] *Id.* ("Decisions referred to in subsection 1 shall not be based on special categories of personal data unless suitable measures to safeguard the data subject's legally protected and legitimate interests are in place.").

[205] *Id.* ("Profiling that results in discrimination against natural persons on the basis of special categories of personal data shall be prohibited.").

on automated processing in criminal matters. Consequently, the risks I pointed out previously could occur.

ii. Ireland

The solution is similar in Ireland. Chapter 4, Section 89 of the Data Protection Act (2018) concerns rights in relation to automated decision-making.[206] Paragraph 2 states a principle of authorization.[207] Paragraph 3 adds a prohibition of discrimination in the case of profiling.[208]

iii. United Kingdom

The United Kingdom decided to follow a similar solution for both civil and criminal matters.[209] Section 49 of the Data Protection Act concerns the right not to be subject to automated decision-making.[210] The UK's criminal, civil, and commercial laws

[206] Data Protection Act 2018 (Act No. 7/2018), ch. 4, sec. 89 (Ir.).

[207] *Id.* The Act allows for automated decision-making when:

> (a) the taking of a decision based solely on automated processing is authorized by the law of the European Union or the law of the State and the law so authorising contains appropriate safeguards for the rights and freedoms of the data subject, including the right of the data subject to make representations to the controller in relation to the decision, and
>
> (b) the controller has taken adequate steps to safeguard the legitimate interests of the data subject.

Id.

[208] *Id.* ("Profiling that results in discrimination against an individual on the basis of a special category of personal data shall be prohibited.").

[209] *See* Data Protection Act 2018, ch. 12, sec. 49 (UK). The data controller is authorized to make a significant decision based solely on automated processing if he guarantees the data subject a right of information and a right to reconsider the decision based on automated processing. *Id.*

[210] *Id.* at cl. 49. According to paragraph 1, "A controller may not take a significant decision based solely on automated processing unless that decision is required or authori[z]ed by law." *Id.* Clause 50, paragraph 2 adds some safeguards:

> Where a controller takes a qualifying significant decision in relation to a data subject based solely on automated processing—
>
> (a) the controller must, as soon as reasonably practicable, notify the data subject in writing that a decision has been taken based solely on automated processing, and
>
> > (b) the data subject may, before the end of the period of 1 month beginning with receipt of the notification, request the controller to—
> >
> > (i) reconsider the decision, or

all cite to exactly the same provisions, even though the risks may be higher in a criminal matter and the data subjects expect more safeguards.

As a procedural matter, the UK applies the GDPR and implements the Directive. However, it preserves its own way, and the ECJ has no jurisdiction to apply a judicial oversight and impose its interpretation. Consequently, it is less relevant for the UK than for the other member states to wonder whether the implementation perfectly respects the European law. Nevertheless, the effects of Brexit are not yet well understood.

Altogether, the European rules are not only too weak but also too diverse, thanks to the enactment of an EU Regulation instead of a Directive. A Digital Single Market Strategy was adopted on May 6, 2015 and was built on three pillars: (1) better access for consumers and businesses to digital goods and services across Europe; (2) creating the right conditions and a level playing field for digital networks and innovative services to flourish; and (3) maximizing the growth potential of the digital economy.[211] Barriers result in citizens missing out on goods and services. Nevertheless, the unification of the rule concerning automated decision-making is only partial and, consequently, insufficient. This fact is problematic, especially for the private players who need to base their activities on a single rule inside the digital single market. Despite the digital single market being one of the EU Commission President Juncker's political priorities, the GDPR partially fails to achieve these goals. Additionally, the consequences of this regulatory failure are potentially catastrophic, as economic activities are increasingly based on algorithmic processing, and the technological potentials are enormous. The risk is to create different levels of protection and requirements inside the EU to regulate such tools, thus resulting in different levels of competition between the member states.

> (ii) take a new decision that is not based solely on automated processing.

Id. at cl. 50.

[211] *See Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Single Market Strategy for Europe*, at 3, COM (2015) 192 final (May 6, 2015).

IV. SOLUTIONS

The algorithmic system has to be improved and not eliminated. Other means for addressing the problem of biases and discrimination are needed. I propose some recommendations, which I hope propel discussion moving forward.

First, some recommendations address the algorithmic system itself. The most important improvement should be sharing information on the system's existence, the motivation for using it, and its goal. Indeed, when data subjects can access intentions, they can better understand and challenge such algorithmic systems.

Moreover, it will be useful to explain which data went into the model (i.e., inputs) and why. Revealing these sources gives residents the opportunity to identify potential bias from data impregnated by historically discriminatory practices. Furthermore, describing how developers analyzed the data could also be a requirement without asking for publishing source code. Access to this information may allow the public to know how developers get from data to output. Additionally, the publication of the performance data creates knowledge as to whether the policy goals initially communicated are achieved.

Some restrictions related to such purposes could be requested. One also may prohibit certain kinds of algorithms if an explanation cannot be given to the individuals requesting it. Such a framework is suggested by the French *Conseil constitutionnel* to encourage the use of algorithms able to satisfy the transparency and accountability requirements.[212] This excludes the use of machine learning methods with the ability to improve their performance by themselves, as well as algorithms protected by secrets and IP rights. More broadly, we could encourage transparency and accountability for both the government and the private data controllers making the administrative decisions. The explanation of the algorithms' characteristics used to make governmental decisions could be extended for all kind of decisions, without consideration of the private or public sector.

[212] *See* Conseil Constitutionnel [CC] [Constitutional Court] decision No. 2018-765DC, June 12, 2018, J.O. 141, recitals 66, 70, 71, at 12–13.

Second, other requirements could concern the ability to audit the algorithmic system, for instance, by scholars or public regulators. The results produced by machine learning systems are best checked for bias and discrimination risks through an audit. Of course, the audit has to respect the guarantees of professional trade secrets. Moreover, it is not mandatory to have access to the source code to control it.

Third, we have to consider the opportunity to establish a powerful regulator with a broad jurisdiction (including consumer and competition law issues) and significant capabilities.[213] An administrative remedy with strong penalties is also necessary.

Fourth, there is opportunity to question when specific and explicit rules of liability should arise, especially in determining whether the human data controller is liable, without consideration of the outcomes generated by the algorithms.

Finally, such proposed rules for algorithmic decision-making are not necessarily related to the processing of personal data. Consequently, it is better to separate them from personal data regulations and to enact specific laws for this specialized area.

CONCLUSION

To sum up, I have shown that the European framework fails to address the discrimination and opacity problems of the algorithms related to machine learning processing and fails to provide a right of explanation regarding outcomes of automated decision-making. The goal of algorithmic transparency is not yet successfully ensured in the EU. The Member States could remedy this by giving more guarantees in this regard, such as the French law seems to do. However, this would create another problem, as companies and data subjects would have difficulty navigating the diversity of rules enacted by the Member States regarding algorithms and automated decision-making. Finally, I propose some recommendations for improving the awareness of and accountability for algorithmic and automated decision-making.

[213] *See* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 117 (2017).