Ethics and Information Technology (2022) 24:17
https://doi.org/10.1007/s10676-022-09642-1

ORIGINAL PAPER

Algorithmic decision-making employing profiling: will trade secrecy protection render the right to explanation toothless?

Paul B. de Laat[1]

Accepted: 9 February 2022 / Published online: 5 April 2022

Abstract
Algorithmic decision-making based on profiling may significantly affect people's destinies. As a rule, however, explanations for such decisions are lacking. What are the chances for a "right to explanation" to be realized soon? After an exploration of the regulatory efforts that are currently pushing for such a right it is concluded that, at the moment, the GDPR stands out as the main force to be reckoned with. In cases of profiling, data subjects are granted the right to receive meaningful information about the functionality of the system in use; for fully automated profiling decisions even an explanation has to be given. However, the trade secrets and intellectual property rights (IPRs) involved must be respected as well. These conflicting rights must be balanced against each other; what will be the outcome? Looking back to 1995, when a similar kind of balancing had been decreed in Europe concerning the right of access (DPD), Wachter et al. (2017) find that according to judicial opinion only generalities of the algorithm had to be disclosed, not specific details. This hardly augurs well for a future right of access let alone to explanation. Thereupon the landscape of IPRs for machine learning (ML) is analysed. Spurred by new USPTO guidelines that clarify when inventions are eligible to be patented, the number of patent applications in the US related to ML in general, and to "predictive analytics" in particular, has soared since 2010—and Europe has followed. I conjecture that in such a climate of intensified protection of intellectual property, companies may legitimately claim that the more their application combines several ML assets that, in addition, are useful in multiple sectors, the more value is at stake when confronted with a call for explanation by data subjects. Consequently, the right to explanation may be severely crippled.

Keywords Copyright · DPD · GDPR · Machine learning · Patent · Patent value · Profiling · Right to explanation · Trade secrecy

Introduction

In this age of machine learning (ML) and artificial intelligence (AI), decision-making based on suitable algorithms is all around us. A recent EU report provides a useful summary of the wide range of applications involved: search engines, filtering of spam or malware, news aggregators, algorithmic journalism, targeted advertising, product recommendation, personalized pricing, and profiling and scoring applications (algo:aware, 2018, pp. iii, 10–11).[1] While all of them deserve societal scrutiny, profiling/scoring applications that significantly affect people's life chances, choices, and opportunities deserve it in particular. Such identification of behavioural patterns based on processing personal data is used by insurance companies, banks, tax departments, police, security forces, schools, and public authorities generally. In those instances, irrespective of whether public or private institutions are involved, full accountability to the members of society should urgently be put on the agenda.[2]

Much has been written about the desiderata on this agenda: algorithms are to be fair and equitable, transparent and explainable, robust and resilient, privacy-proof, and accountable (cf. algo:aware, 2018, pp. iv, 5–6). Below I will interchangeably use the roughly equivalent terms "responsible AI" (in use among companies; cf. de Laat, 2021) and "trustworthy AI" (coined in EU circles) for AI that conforms

✉ Paul B. de Laat
p.b.de.laat@cerug.nl

[1] University of Groningen, Groningen, The Netherlands

[1] Although profiling has a broader meaning than scoring, the two terms will be used interchangeably.

[2] A recent report by AlgorithmWatch has a similar focus on automated decision-making systems, based on algorithms, in so far as they affect "justice, equality, participation, and public welfare" (AlgorithmWatch, 2019, p. 9).

to these principles. On closer inspection, transparency had better be interpreted as the necessary condition for the other desiderata to be realized. After investigating how transparency can contribute to enhancing accountability, I concluded elsewhere that, in particular, individuals are to be entitled to obtain an explanation about algorithmic decisions that affect them. Such explanations may refer to the logic in general of the system involved ("weak explanation"), to the reasons for a specific decision ("strong explanation"), or preferably to both (de Laat, 2018).[3,4]

At the moment such particular transparency is lacking almost completely. Institutions, whether public or private, refuse to provide anything other than trivial details.[5] The well-known FICO scorecards for creditworthiness, as the exception to this rule, are the best information one can get (https://www.myfico.com/credit-education/whats-in-your-credit-score). There are several reasons for this refusal (more details in de Laat, 2018). First, as far as firms are involved, they want to keep their algorithmic recipes a secret. The intellectual resources they invested a lot of money in, are not to leak to their competitors. Secondly, opening up algorithmic details may in some cases enable data subjects to game the system. That is, some knowledge of the proxies involved allows decision subjects to evade them. Thirdly, most of the ML methods currently in use produce opaque models; these cannot be interpreted easily. Derivation of reasons for individual outcomes is simply impossible then—unless considerable efforts are undertaken to wrench some clarity from the models involved.

Given these obstacles, how feasible is it that the provision of full explanations about profiling decisions will become the legal norm in the near future? This is the question to be answered by this study. After a description and assessment of current initiatives of (self-)regulation I analyse whether the General Data Protection Regulation (henceforth: GDPR) may contribute to data controllers providing explanations about their algorithmic decisions to data subjects as far as Europe is concerned. Trade secrecy is identified as the critical factor that may decide whether or not institutions can be forced to open up. The crux is found to lie in the legally required balancing of access/explanation rights of data subjects against trade secrecy and IP rights of data controllers. Subsequently I explore the landscape of intellectual property protection and show that patenting related to ML has sharply increased since 2010. I argue that as a result the legal protection usually granted to trade secrets concerning algorithmic processing may in certain instances acquire a quasi-absolute status, effectively annihilating the prospect of a legally effective right to explanation emanating from the GDPR. In the final section it is tentatively suggested that other demands for explainable AI, in particular from the US military, may save the situation.

Explainability

Providing information about the logic involved in profiling surely is a technically possibility. But what about providing reasons for specific algorithmic decisions based on profiling? Are "strong explanations" of the kind feasible at all? Since ML has become the basic tool underlying all modern-day profiling, the urgent question imposes itself whether models developed by ML can readily be explained (in the strong sense). Let me therefore first, before the ensuing legal analysis, present a short review of the state of the art concerning this issue of "explainability". Until recently, the main trend in ML was towards increasing sophistication in pursuit of maximal accuracy (de Laat, 2018, Sect. 7, p. 536 ff.). Think of bagging and boosting of classifiers that result in a dense forest of trees that must be summated (cf. more details and references in de Laat, 2018, p. 537). Similarly, neural networks are being applied with ever more intermediate layers between input and output. All such sophistication obfuscates interpretation.

Recently, multiple authors have opened up new lines of inquiry in order to confront the problem of explainability.[6,7] Techniques can be tailored to a specific ML-model (say an ensemble of trees), or, more generally, be applicable to any ML-model imaginable (model-agnostic approach). To begin with, textual or visual explanations may be provided, often in combination with other techniques. Further, "feature relevance" explanations try to reveal the influence of various model features on its output. "Sensitivity-based

[3] In that publication I focussed on transparency about algorithmic decision-making *as a whole*. Here, I want to focus specifically on clarification of the last phase in which algorithmic decisions are made.

[4] Cf. below (section on GDPR) for more details about my position in this regard.

[5] Frank Pasquale was one of the first scholars to draw attention to this lack of transparency. For automated predictions such as credit scoring cf. Citron & Pasquale (2014); for a more general overview cf. his 2015 book (Pasquale, 2015).

[6] This overview is based on the following sources. A useful classification of techniques currently in use is to be found in the exhaustive overview (over 400 references) provided by Arrieta et al., (2019; cf. in particular par. 2.5.2). Further, Molnar (2021) provides an overview of the mathematical foundations of the tools involved, while de Laat (2021) gives an overview of the various software implementations of them. More general sources are algo:aware (2018, pp. 25–26), DARPA (2016, pp. 7–8), Edwards and Veale (2017), and Lipton (2016). A critique of the approaches mentioned is given in Mittelstadt et al. (2019).

[7] Notably, these efforts largely came about in reaction to the publication of the GDPR proposal in 2016.

explanations" and the method of "counterfactuals" (Wachter et al., 2018) belong to this category. Next, "explanations by example" try to catch representative examples of the set of input data that grasp the essence of the model developed (aka case-based explanations). With the technique of "local explanations", the ML practitioner focusses on a restricted subset of input data and tries to explain how the model classifies them (aka demographic-based explanations). Finally, "explanation by simplification" denotes learning a new, simpler model; it is supposed to mimic the behaviour of the original, more complex model. Model-agnostic LIME, focussing locally on an area around a particular data point, is a famous example of this technique (Ribeiro et al., 2016).

The *post-hoc* methods just mentioned are needed when models are non-interpretable ("black boxes"). However, one may do away with complex models and restrict one's ML to generating simpler models that can be interpreted *by design* (such models are referred to as "interpretable" models). Some examples are Bayesian rule lists (Letham et al., 2015) and generalized additive models (Caruana et al., 2015; Lou et al., 2012). Whether or not accuracy necessarily has to be sacrificed in the process is a hotly debated issue (Rudin, 2018).

It has to be stressed that currently these novel techniques are not in wide-spread use. A recent McKinsey study among firms using AI found that only 19% of them are "actively addressing" the risk associated with explainability (among "AI high performers" the percentage rose to 42%) (McKinsey, 2019; cf. exhibit 4).[8] Moreover, the techniques for explainable ML in organisations generally are mainly used as yet by ML engineers internally as "sanity checks" on their models: while often inconsistencies emerge between a model's outcomes and the intuition of these engineers (or of the domain experts involved), techniques focussing on explanation may help to resolve them (Bhatt et al., 2019). From these observations it becomes clear where a legal right to explanation may usefully come in. It may have a *dual* function: fostering the adoption of explainability tools by a much wider audience of AI using organisations generally and stimulating the actual transfer of such techniques from the ML "laboratory" to organisational departments tasked with providing explanations to end users. Obviously, both developments build upon each other.

[8] When the investigation was repeated in 2020, the percentage among AI using firms had increased slightly to 25%. In 2021, for developed economies at least, the percentage was higher again: 30%.

Regulation

After this brief overview of research approaches towards the explainability problem I return to my main question: will the forces of regulation succeed in establishing a right to explanation that is legally effective? An affirmative answer to this question is evidently important since organisations subjected to regulation may as a result feel forced to pick up the research clues about explainability and incorporate the new techniques in their algorithmic repertoire.

To that end, let me first chart the relevant forces of regulation currently in operation, with a focus on the American and European continents. In the US, several Congress resolutions are under study (cf. overview at https://futureoflife.org/ai-policy-united-states/)). House Resolution 153 (2019), introduced in 2019, supports the development of ethical guidelines for AI. Further, the "Algorithmic Accountability Act" (2019), also introduced in 2019, requires organisations that process personal data to conduct impact assessments related to automated decision making and data protection. At the local level, the city of New York, in 2017, installed a task force to study the use of automated decision systems by the city's agencies. It came up (in 2019) with a report that recommends setting up a central agency for coordination and management of such systems. Guidelines for algorithmic decision-making, along principles of responsible AI, are to be developed (New York City Automated Decision Systems Task Force, 2019). At state level, the "California Consumer Privacy Act" (CCPA) (2018), which passed in 2018 and came into effect January 2020, stands out as the major American development concerning regulation of AI. Modelled after the GDPR (though less stringent on data controllers), it is all about the protection of consumer data. These initiatives are ongoing still—except for the CCPA that has passed. Returning to the major focus of this study: does explainability figure in any of them? Both House Resolution 153 and the report from the New York task force indeed contain a recommendation to provide explanations about decisions to end users.

In neighbouring Canada, several initiatives that touch on the issue of explanation of algorithmic decisions are also evolving simultaneously. At the federal level bill C-11 (2020), which was introduced in 2020, focusses on privacy and data protection concerning commercial activities. It requires organisations that use automated decision systems for predictive purposes to provide the data subject with "an explanation of the prediction, recommendation or decision" (par. 63.3). Further, a discussion is underway about a proposal for reforming PIPEDA, the federal privacy legislation that (mainly) regulates private sector firms as to their handling of personal data. Strongly emulating the GDPR, it proposes a "right to meaningful explanation" of automated

decisions (OPC, 2020).[9] At state level, Québec in September 2021 accepted Bill 64, an act that modernizes other acts as regards the protection of personal information (comes into force 2023) (Bill64, 2021). It requires organisations, *both* public and private, involved in *fully* automated processing of personal data to inform data subjects about "the reasons and the principal factors and parameters that led to a decision" (pages 15, 37).

In Europe, state regulation is more prominent, with the GDPR (2016; in force from 25 May 2018) being the prime example of a regulation impacting data protection in general, and transparency in particular. Before embarking on an extended discussion of this GDPR below, it should be noted that the EU has already set the next step: discussions are staged at several levels about the future of our AI-society. For example, the "EU high-level expert group on AI" has proposed guidelines for "trustworthy AI" (EU, 2019). This refers to AI that is ethically *and* technically sound. In the guidelines explainability of AI figures as one of the defining features of trustworthy AI. In the same vein, a slew of other EU agencies and committees, as well as the EU parliament, have performed studies and drafted recommendations about the issues of robotics and AI (cf. AlgorithmWatch, 2019, pp. 19–25).

Subsequently, the European Commission has opened up the discussion about an appropriate regulatory framework for AI that will create an "ecosystem of trust" (EU, 2020). Its "White Paper about the future of AI in Europe" focusses on AI systems that pose high risk. As regulatory implementation the Commission recently (April 2021) formulated its proposal for an "Artificial Intelligence Act" to be discussed in the ensuing months (EU, 2021). The Act requires high-risk systems to fulfil the criteria for trustworthy AI; explainability in particular. Using the more general term transparency, it states that "for natural persons, a certain degree of transparency should be required for high-risk AI systems. Users should be able to interpret the system output and use it appropriately" (EU, 2021, recital 47, p. 30).[10]

In addition to such state regulation at the EU-level, many self-regulatory initiatives are unfolding that (may) relate to a right to explanation (for a complete overview see algo:aware, 2018, pp. 37–109). For a start, various standard setting bodies have begun to focus on algorithms. Organisations such as BSI, CEN/CENELEC, ETSI, IEC, and ISO develop standards, benchmarks, guidelines, codes of conduct, and the like. Moreover, they can certify organisations as having adopted those very standards. The EU has even explicitly invited some of them to do so concerning the GDPR (GDPR, article 42 on certification). Picking up the clue from the EU high level expert group, several standardization bodies are discussing the theme of trustworthy computing. A first standard was delivered mid-2019: ISO/IEC 27701, focussing on "privacy management", aka protection of personal data. Since GDPR obligations have been mapped onto requirements in the new standard, certification for 27,701 promises to demonstrate evidence of compliance with the GDPR. On closer inspection, no specific requirements on data controllers are mentioned in the ISO/IEC standard concerning explanations of algorithmic decisions. The precise terms specifying such rights in the GDPR have disappeared completely.[11] For some reason, nothing of those putative rights survived the mapping exercise.

Further, companies have definitely woken up to the challenge of subscribing to principles for "responsible AI". In 2016 the main high-tech companies (Amazon, Apple, Facebook, Google, IBM, and Microsoft) initiated the "Partnership on AI" platform. It explicitly mentions a focus on "fair, explainable, and accountable" AI systems (www.partnershiponai.org). From 2016 onwards, some 25 mostly large companies have pledged to adhere to ethical principles for responsible (or trustworthy) AI. What is more, some of them have developed practical software tools for such AI. To wit, as far as producing explanations for algorithmic decisions is concerned: Google (What-If tool), Microsoft (InterpretML), and IBM (AI Explainability 360 Toolkit). All these packages have been made available as open source. This corporate move as a whole towards responsible AI has extensively been charted by de Laat (2021).

Finally, professional organisations have thrown themselves into the discussion. The IEEE is currently developing a series of standards for ethical AI. Its "Ethics Certification Program for Intelligent and Autonomous Systems" will develop standards for the accountability, transparency, and reduction of algorithmic bias of such systems. Further, its P7000 series develops standards for taking ethical concerns

[9] By way of clarification the report states that "the right would be similar to what is found in Article 15(1)(h) of the GDPR." That remark unfortunately muddles the distinction between the functionality of a profiling system and the explanation of a specific decision (cf. below).

[10] Interestingly, in view of the central theme of this research, the proposal in one passage appears to take a stand on the issue of balancing rights of IP protection vs rights of data protection: "The increased transparency obligations will also not disproportionately affect the right to protection of intellectual property (Article 17(2)), since they will be limited only to the minimum necessary information for individuals to exercise their right to an effective remedy (..). Any disclosure of information will be carried out in compliance with relevant legislation in the field, including the [Trade Secrets Directive of 2016]" (EU, 2021, par. 3.5, p. 11).

[11] The precise wording that is missing: "meaningful information about the logic involved (..)" in articles 13–15 (right of access; weak explanation), and "an explanation of the decision reached" in recital 71 (right to explanation; strong explanation); cf. analysis below.

into account while developing such systems; its focus is on robotic systems. In order to receive inputs and inspiration, the IEEE published a report entitled "Ethically Aligned Design" (IEEE, 2018). This broad discussion piece specifically mentions the right to explanation whenever automatic decisions are taken (IEEE, 2018, pp. 152, 160)—citing European developments in law as prominent example.

Will these regulatory initiatives impact the right to explanation? And if so, will the technical methods and tools for explanations to users that meanwhile do exist be effectively incorporated into corporate and organisational practices for profiling/scoring as a result of the forces of regulation? It can be concluded from the above that, for the moment, only the GDPR has real force—all other relevant laws and standards are still in the discussion stage. Therefore, it seems safe to state that at least for the near future it is only that regulation that can effectively have an impact on realizing a legal right to explanation for European citizens. Let me proceed from here and investigate the connections between the GDPR and one or other right to explanation.

General Data Protection Regulation

The GDPR (EU Regulation, 2016/679) is intended for the protection of personal data.[12] Many commentators claim, however, that it may also provide protection beyond the immediate data themselves: concerning the upstream *processes* that use the personal data, read profiling or algorithms in general. This Regulation will now be subjected to a detailed analysis concerning the question whether, and to what extent, it grants data subjects the right to obtain an explanation of how decisions came about. In the process I rely on the GDPR itself as well as the guidelines developed by the Article 29 Data Protection Working Party (referred to as A29WP),[13] and on several juridical commentators.

The term explanation will be used in a wide sense: *all* pieces of information that may be more or less informative for a data subject for understanding how a decision came about will be subsumed under the term "explanation". This may range from information about the variables in use, their weights, the scoring formula, or the ML model trained—together to be referred to as "weak explanation"—to information about the reasons for a decision as produced by some explanatory method (such as local explainability, counterfactual explanation, explanation via examples, or interpretable algorithm)—to be referred to as "strong explanation".[14] Obviously, what data subjects want most of the time, is a strong explanation which allows them to see the grounds for a decision about them and, if needed, contest it. When in this text I refer henceforth to the right to explanation without any further clarification, as a rule I mean the strong kind of explanation.[15]

For systems processing personal data (including profiling) that have significant effects on people's lives, several routes to a "right to explanation" of decisions can be discerned, via respectively articles 13–14-15 and recital 71 (cf. Wachter et al., 2017). To begin with, data controllers have a duty to notify data subjects whenever their personal data are being processed (articles 13, 14). If profiling is involved, they must specify "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" (article 13.2.f; article 14.2.g). The other route is through article 15: the right of access. Data subjects have the right to know whether their personal data are being processed; if this occurs by means of profiling, they are entitled—again—to "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" (article 15.1.h).

How are these sentences to be interpreted? The A29WP has provided further guidelines. Meaningful information about the logic involved does not necessarily imply "a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision" (A29WP, p. 25). As for significance and envisaged consequences, the guidelines clarify that "the controller should provide the data subject with information about the *envisaged consequences* of the processing, rather than an explanation of a *particular* decision" (A29WP, p. 27; italics in original). The A29WP specifies that "real, tangible examples of the type of possible effects should be given;" the example provided is an app that compares insurance premiums for dangerous vs safe drivers (A29WP, p. 26). In sum, according to the interpretation of

[12] The EU adopted a separate data protection directive for the spheres of police and criminal justice authorities (EU Directive 2016/680). Similar provisions as in the GDPR apply—though with more restrictions on the right of access. This directive is not considered here.

[13] Interpretations of the A29WP, consisting of (representatives of) Data Protection Authorities, have substantial weight in the juridical debate ("soft law"). It has been succeeded by the European Data Protection Board in 2018.

[14] Cf. section above on Explainability.

[15] In anticipation of the discussion that follows: in my interpretation of the GDPR, articles 13–15 correspond to a weak explanation for decisions, while (parts of) recital 71 correspond to a strong explanation. Selbst and Powles (2017), to be discussed below, take the same position: they extensively argue, for example, that the right to obtain "meaningful information about the logic involved" (articles 13–15) also constitutes a right to explanation. In contrast, Wachter et al. (2017) designate this as (only) the "right to be informed".

the A29WP—which squares effortlessly with the analysis by Wachter et al. (2017)—the disclosures of articles 13–15 clearly refer to functional information about the profiling process (*ex ante*), but not to details about how specific decisions have been reached—let alone an explanation of them (*ex post*).[16]

Further, systems processing personal data that significantly affect people can, ultimately, take decisions in fully automatic fashion—humans are no longer involved. Though these are normally forbidden (article 22.1), many exceptions apply—for reasons of contract, fraud prevention, and the like such systems are allowed. Whenever they are used, *additional* safeguards for data subjects must be granted: "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" (article 22.3). Surprisingly, recital 71 formulates yet another safeguard: data subjects acquire the right "to obtain an explanation of the decision reached" (recital 71). So here we observe the GDPR for the first—and only—time mentioning that subjects have some right to *explanation* of decisions—be it only for *fully automated* systems. Not unimportantly, the A29WP, on its part, has explicitly endorsed this requirement, arguing that "the data subject will only be able to challenge a decision or express their view if they fully understand how it has been made (..)" (A29WP, p. 27).

However, as every commentator of the GDPR hastens to add, recitals do not have the same status as articles in a regulation. Articles are rules of law; recitals are just explanatory comments on them.[17] Note as well, that inserting some (nominal) human steps into the decision-making seems to provide an easy loophole to circumvent all the safeguards mentioned; then, only articles 13–15 must be respected. But, as the A29WP warns, oversight of the decision should be "meaningful, rather than just a token gesture" (A29WP, p. 21). In sum, this recital (71) seems to represent some—though fragile—basis for a right to explanation concerning the subclass of fully automated (profiling) decisions.

So, at first sight the sketched routes to obtain explanations look promising: information about the *functionality* of the profiling system (as Wachter et al., 2017 phrase this) has to be disclosed, and additionally an *explanation proper* seems to be required whenever profiling occurs in fully automated fashion. However, there is also recital 63 to be taken into account, which explicitly introduces restrictions to the right of access: "That right [of access] should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software." It can reasonably be assumed that these restrictions not only apply to the right of access (articles 13–15) but also to the fragile right to obtain an explanation (recital 71). After all, elements of an explanation readily reveal more about underlying algorithms than functional accounts about them. Noticeably, although recital 63 is only a recital, its legal status cannot be considered to be fragile, while it has all the backing of a separate EU law, the Trade Secrets Directive (to be abbreviated as TSD; EU Directive, 2016/943). This TSD, implemented in national law in most European countries from 2018 onwards, has clarified and fortified the legal status of trade secrets.

Let me, before embarking on a discussion of this process of balancing, summarize my findings about the GDPR and purported rights of access/explanation. When profiling occurs, data subjects may require information about the *functionality* of the system in use (articles 13–15); if the profiling proceeds in fully automated fashion, three additional safeguards apply (article 22); a fourth one is the right to obtain an *explanation* (recital 71). Such requests by data subjects, though, always have to be balanced against the protection of the trade secrets and IPRs involved (recital 63). Not unimportantly, that protection also has the full force behind it of a separate law, the TSD. The outcome of this balancing act remains an open question.

Balancing of interests in the GDPR in comparison with the DPD: Wachter et al.

The GDPR therefore provides data subjects with rights of access and explanation that have to be balanced against the claims for protection of trade secrecy and IPRs as far as these are put forward by the proprietors of the algorithms involved. In order to assess how this balancing of interests may work out in the future, Wachter et al. (2017) engage in an intriguing comparison. They look back to 1995, the year that the Data Protection Directive—the precursor of the GDPR—came into force. Already then, data subjects were given a right of access whenever their personal data were being processed. They were entitled to obtain "knowledge of the logic involved in any automatic processing of data concerning him," at least concerning *fully automated* decisions (EU Directive 95/46/EC: article 12.a).[18] At the same time it was stipulated that "this right [of access] must not adversely affect trade secrets or intellectual property and in particular the copyright protecting the software" (idem: recital 41).

[16] The A29WP interpretation relies on the same distinctions as introduced by Wachter et al. (2017): information about system functionality vs information about a particular decision, and information *ex ante* vs information *ex post*.

[17] As Wachter et al. (2017) ironically note, in earlier drafts of the regulation the explanation provision sat squarely in article 22. After lengthy consultations and negotiations, it was moved to secondary status in a recital.

[18] But notice: *not* to an explanation of specific decisions.

So, a balancing procedure was already indicated from 1995 onwards—remarkably similar in wording to the GDPR.

How did the balancing work out in practice from 1995 on in European law and European courts (since the Directive had to be implemented in national laws)? Wachter et al., (2017, pp. 85–90) devote many pages to these developments; let me render their conclusion in succinct fashion. The middle ground that gradually emerged from this balancing consisted of opening up some generalities but carefully hiding concrete details. Some examples: the logic of a decision tree, but not its parameters that decide the precise outcome; features in use, but not the specific weights attached to them. Opening up a scoring formula, an algorithm, let alone pseudo code or source code are, obviously, totally out of the question—let alone, in these pre-GDPR times, any explanation along the lines sketched above (in the section on explainability). This conclusion applies to European countries such as Austria, France, Germany, and the UK. Jurisprudence, legal commentaries, and academic literature converge on this interpretation of the 1995 Directive's right of access. Two factors are considered legitimate grounds for keeping algorithmic information exempt from disclosure. Sporadically it is argued that disclosing such details would invite "gaming" by *users* (Wachter et al., 2017, note 84) and therefore prejudice the commercial interests involved. More often, though, the argument is that trade secrets or IPRs (copyright in particular) are to be protected. Revealing details could lead to the desired secrecy being broken, or the copyrights involved being infringed. That is, *competitors* may profit from disclosures.

This analysis of the scope of the right of access from 1995 onwards does not augur well for a right of access let alone a right to explanation emanating from the GDPR in the near future. After all, the juridical base line for the mandatory balancing of interests hardly changed from the 1995 DPD to the 2016 GDPR. In a subsequent study, Wachter and Mittelstadt (2019) raise even more warning flags. According to them, recent jurisprudence[19] of the Court of Justice of the EU (CJEU) concerning data protection indicates that "personal data" should be interpreted in a narrow sense; only data directly linked to the person are to count as such, not any inferences (like profiles) drawn from them (Wachter & Mittelstadt, 2019, part IV). Moreover, according to the Court, the purpose of data protection law is not to assess the accuracy of decision-making processes involving personal data. In view of these two considerations, Wachter and Mittelstadt (2019, part IV) conclude that the right of access (article 15 of the GDPR) thus becomes significantly curtailed**.**

Expectations about the right to explanation in the GDPR: other scholars

Before proceeding let me mention some comments by other scholars on the central issue of balancing conflicting interests. Some are just as pessimistic as Wachter et al. (2017). After a close reading of the GDPR and the TSD Gunst (2017) argues that from a legal point of view, one cannot conclude that either trade secrecy protection or data protection prevails; their balancing therefore must be decided on a case-by-case basis. Turning to case law for guidance, she found that the CJEU has never adjudicated cases involving that precise balance—but it has treated cases in which trade secrecy considerations had to be balanced against *other* legitimate interests. The Court invariably ruled that such conflicts had to be resolved along the principles of fairness and proportionality—and referred them back to the national courts. This underscores the importance of case law in various European countries in the DPD-era about the balancing issue—which clearly shows a pattern of trade secrets having a strong limiting effect on data subjects' access rights (as similarly concluded by Wachter et al., 2017). She concludes that the prospects for a right to explanation to materialize "must be viewed with [a] certain disbelief" (Gunst, 2017, p. 84).

Similarly, after observing that in the era of the DPD German courts have consistently ruled that a description of the abstract design of the system was generally sufficient as accounting for the logic involved in fully automated decisions (Wischmeyer, 2019, para 16), Wischmeyer concludes that, as the legal grounds for secrecy continue to exist much the same, "there is little ground to assume that this jurisprudence will change significantly under the GDPR" (Wischmeyer, 2019, para 17).

Other scholars appear to be more optimistic about a right to obtain explanations. After a legal analysis of the right of access vs the right to trade secrecy in both the GDPR and the TSD, Malgieri (2016)—in line with Gunst (2017)—concludes that none prevails a priori; therefore, balancing on a case-by-case basis is indicated. In order to resolve the apparent conflict, he proposes the method of "de-contextualization": taking the data in dispute out of the economic context (Malgieri, 2016, pp. 105, 114, 115). By way of example, he mentions a company that has generated customer profiles; the right of access, Malgieri claims, can easily be satisfied by granting clients access to their own personal data—but not to lists of clients, profiles, forecasts, and the like. In this way, presumably, the rights in conflict are both satisfied. To me, this novel "solution" has all the appearances of trade

[19] Their study discusses court cases C-141/12 and C-372/12 (adjudicated during 2012–2014), and case C-434/16 (adjudicated during 2016–2018).

secrecy gaining the upper hand; the right of access is served in just a minimalist fashion.

In a subsequent article, Malgieri (with Comandé) has become more optimistic about the required balancing act itself (Malgieri & Comandé, 2017). This time they argue that the GDPR can be interpreted as exhibiting a legal preference for data protection rights *over* trade secret protection rights. Their main argument is that the GDPR states that data protection rights "should not adversely affect" trade secrets (recital 63), whereas the TSD merely states that trade secret discipline "should not affect" data protection rights (recital 35) (Malgieri & Comandé, 2017, pp. 263, 264). Apart from the observation that I cannot imagine any "affect" that is not an "adverse affect"—which would eliminate any purported preference—this glosses over the fact that whenever these rights seriously clash in practice, finding a case-by-case solution is usually recommended that does fair and proportional justice to both interests. As Gunst (2017) has shown, the CJEU has always refrained from affirming the primacy of any of those rights; let alone suggesting a precise calculus for trading off between those interests. So, I am not really convinced that the very process of balancing recommended in the GDPR is slanted towards data protection rights.

Lastly, I want to mention several publications by American scholars that emanate positivity about the right to explanation, though often from considerations *other* than the required balancing of interests. Let me first mention Selbst and Powles (2017) who vehemently oppose Wachter et al. (2017). First and foremost, the two authors attack the very distinctions made by Wachter et al. (system functionality vs specific decision; *ex ante* vs *ex post*) as artificial and distorting the discussion. Moreover, they qualify the historical analysis by Wachter et al. (2017) of the last two decades of DPD as "not conclusive" while only involving "non-binding interpretations" of national courts. Finally, they argue that the GDPR—in comparison to the DPD—introduces new elements that may contribute to more transparency about profiling for data subjects: not just "knowledge" of but "meaningful information" about the logic involved has to be provided (articles 13–15), and both transparent processing of personal data (article 5) and the willing cooperation of data controllers when data subjects exercise their rights (articles 12, 22) are emphasized. Based on this "textual optimism" they ultimately conclude that articles 13–15 do indeed constitute some right to explanation.

I have no issue with that modest conclusion (after all, a weak explanation *is* a contribution to understanding a decision; cf. my definition of "explanation" above), nor with their "textual optimism" about the GDPR underlying that conclusion. But I do take issue with the other points mentioned. Dismissing national court rulings as "inconclusive" fails to appreciate that such verdicts do carry weight, since, if relevant cases are brought before the CJEU, as a rule these are referred back to the local court(s) in question—that is, if they ever reach that Court at all (cf. Gunst, 2017). Further, more gravely, their attack on the distinctions made by Wachter et al. seems unfounded and based on a distorted picture of ML.[20]

Secondly, Kaminski cannot be ignored. While also opposing in particular the position taken by Wachter et al. (2017), she argues that articles (13–15 & 22), when interpreted in the light of the GDPR recitals and the A29WP guidelines, "put in place an algorithmic accountability regime [which relies on auditing and ethical review boards; PBdL] that is broader, stronger, and deeper than the largely symbolic regime that existed under the DPD" (Kaminski, 2019, p. 208). As to the right to explanation as mentioned in recital 71, she argues that it is not so shallow as some have argued, when—again—considered in the context of other GDPR articles and the A29WP guidelines. Moreover, recital 71 does not have to be debunked as only a recital, since recitals do have argumentative standing in court and may explicitly be considered by Data Protection Authorities (DPAs). These may also be on guard for overly broad trade secrecy claims.

Casey (2018), finally, is also optimistic about the right to explanation. His optimism does not follow from a close reading of any of the rights of access or explanation, but merely from positive expectations of DPAs (as introduced in chapters 6 and 8 of the GDPR) interpreting the Regulation in novel and beneficial ways.

In spite of the arguments advanced by these more optimistic scholars, I cannot help, though, but remain sceptical about the right to explanation acquiring effective legal status with the GDPR in force. To me, Europe's legal past of uneven balancing the right of access against trade secrecy and IPRs casts too large a shadow over the future to be ignored. Moreover, putting one's cards on already overburdened DPAs is too much of an act of faith for me. Regardless of the specific expectations one may hold about the future, though, whether more negative or more positive, I think it is time to shed light on the issue from another perspective. The scholars referred to above fail to fully appreciate that firms are currently more intent than ever on protecting their achievements in ML. This has important consequences for the status and weight of trade secrecy considerations. The next sections are devoted to developing this argument.

[20] Lack of space forbids going deeper into this issue; but interested readers may be referred to sections 3 and 7 of de Laat (2018) that treat the various phases of algorithmic processing. Note, moreover, that the A29WP essentially employed the same distinctions.

Table 1 Overview of legal protection mechanisms that may pertain to intellectual resources involved in successive stages of profiling (further clarifications in the text)

	Data (input)	Database processing	Profiling/scoring by means of ML algorithms	Model (output)
Copyright, Database rights		*	*	*
Trade secrecy	*	*	*	*
Patents		*	*	

Trade secrecy and intellectual property rights

Let me first provide a short survey of the state of the art concerning the legal protection of intellectual resources as far as these are involved in profiling systems, with, at this time, several new EU laws in force. Table 1 gives an—admittedly—cursory overview and needs to be interpreted as follows.

Copyright can be argued to protect most of the intellectual resources involved in profiling. For profiling software, copyright protects the expression of the ideas behind the software, in its literal form (source code, object code). The models produced, say a decision-tree, presumably also enjoy such protection—although the discussion continues (since they can be argued to be created by a machine, not by a human). The new EU Copyright Directive (adopted in 2019) is unlikely to change this picture much, since most "society service providers" carrying out data mining will be exempted from the new copyright regime (which normally would require a license from data subjects to process their personal data). Note that database rights also exist which are separate from but analogous to copyright. As soon as bare data are organized into databases requiring a "substantial investment", their creators may claim database rights. These *sui generis* rights are established in Europe—but not in the US.

Trade secrecy covers *all* phases of profiling. Whatever information a company considers vital for its competitive position and takes efforts to keep within its walls, may be "declared" a trade secret in the legal sense. Even bare data—whether personal data or not—may fall under this rubric. The new EU Trade Secrets Directive that took effect in 2018 (EU Directive, 2016/943) has clarified and made explicit what may count as a trade secret. Databases with client data, algorithms, and outputted models have all been mentioned as suitable candidates (Wachter & Mittelstadt, 2019, VI.D). Note that in the US such a broad conception of trade secrecy has been entertained for long.

Finally, patenting is to be considered. Patents can be granted on product or process inventions that are useful, novel, and non-obvious. Once granted, one has exclusive rights to make, use, or sell the invention—as well as the right to license it to others. A patent protects the idea behind the invention, not its specific implementation. Accordingly, inventive software developed for profiling purposes is eligible for patent protection. Further, inventive systems related to database functionality may also be submitted for patenting ("database patents"). Of course, neither bare data sets nor models as such can be patented. As will be elucidated below, the last decade has witnessed an upsurge in such patenting, especially in relation to ML.

Now, what is the relevance of these forms of legal protection for understanding the GDPR? To what extent are they actually involved in the required balancing against the rights of access/explanation? Trade secrecy (as mentioned in recital 63) is sure to be an important factor, because of its breadth. Any detail of the algorithmic process, from beginning to end, may be treated as a trade secret by an organisation involved. As such, a clash with said rights seems inevitable. Just a request to inspect the personal data collected by a controller can already (partly) be denied on this score, let alone a request for more intricate details of algorithmic processing.

In the GDPR copyright is mentioned in the same breath with trade secrecy as an important balancing factor; they must be "protected" (recital 63). On closer inspection, this concern seems slightly exaggerated. If proprietors of copyrights disclose some details of their processing, it is difficult to imagine that others may usefully infringe their copyright.

When competitors, say, would proceed to employ any disclosed variables to create a model of their own, copyright is simply not infringed. Only in the extreme case, when, say, a *complete* scoring formula or *complete* model is revealed, infringement turns into a realistic possibility since, by definition, both can be copied and put to practical use without further ado. Such an extreme disclosure, however, would go far beyond the rights of access or explanation as granted by the GDPR. Therefore, when discussing the GDPR and the required balancing of rights, protecting copyrights is hardly a factor to be reckoned with.

Lastly, consider the situation that patents are involved. In my opinion, they also play a minor role in the issue of balancing rights against each other. Disclosures about patents involved do not reveal any new information, since patent applications become public after filing; patents therefore do not have to be "protected" against access. No wonder that patents are not explicitly mentioned anywhere in the GDPR. Nevertheless, the realistic scenario has to be considered that revelation of *which* particular patents are involved in the algorithmic processing may invite "inventing around" them (or even infringement of them). This scenario brings us back, though, to trade secrecy as the pivotal issue.

So, I maintain that the upshot of these considerations is that trade secrecy plays the dominant role in the issue of balancing against the rights of access/explanation. Can we somehow gauge its importance? I propose we are able to do so by looking at the *covariation* of patenting and trade secrecy. This requires some explanation. When new research yields inventions, at first these can only be protected by trade secrecy. After some time, these valuable results may become eligible for patenting. Then, a company has to choose whether to continue protection by means of secrecy or to switch to filing patent applications instead. In that more mature stage, *both* trade secrecy and patenting are feasible options to choose between. The appropriate choice between them will depend on many factors, such as the level of obviousness of the invention and the required amount of disclosure of knowledge (filing a patent requires disclosure to some extent, while trade secrecy does not) (Erkal, 2004).

Given this mixed approach, we may reasonably presume that whenever a company is filing for patents, they also have other intellectual assets that are kept secret. Patenting activity is just a tip of the iceberg of all intellectual resources in their portfolio.[21] If so, patent numbers may serve as a rough *indicator* of the number of trade secrets in a company's possession; secrets that cover all most recent research findings as well as some older ones. One might even argue that the number of patents is more likely to underestimate the number of trade secrets being kept, since companies in general often prefer trade secrecy over patenting as a means of protection (EUIPO, 2017; Willoughby, 2013). To an AI/ML portfolio this may apply in particular (cf. Quinn Emanuel, 2020). Further, filing a patent application for a ML invention requires some amount of disclosure. As a result, a firm may be faced with the request to open up the underlying datasets; these may have been costly to collect and to process (Poursoltani, 2020). If this burden is too high, it may well choose the continuation of trade secrecy over patenting.

This patent numbers indicator may then be useful for our foregoing discussion about balancing various rights. With the number of submitted patents rising over the years for many a company (as shown below), we may reasonably conclude that these companies have, in addition, the same amount of secret intellectual resources at their disposal—if not more. Correspondingly, they have legitimate grounds, to be explored below, to argue that more weight than ever before should be attributed to trade secrecy considerations in any dispute which requires a balancing against the rights of access/explanation (or indeed, any other rights involved).

Patents related to machine learning

Let me now elaborate on this argument, focussing on the field of ML, which is the defining field contributing to profiling applications. Its origins go back to the 1990s, when several pioneers, such as Leo Breiman and Yoav Freund, invented the basic techniques such as bagging and boosting that now belong to the standard repertoire of ML techniques. It was an academic backwater at the time. From 2000 onwards, however, the field grew step by step since the advent of Big Data promised to bring new applications within reach. In parallel, ML scientists began to file applications to obtain patents on their inventions. This may sound strange, as such mathematical techniques would seem to lie far away from practical applications. Therefore, a small excursus into patent law seems indicated (the following paragraph is based on de Laat, 2000).

In the 1980s patent offices in the US were reluctant to patent software-related inventions, since abstract ideas, mathematical formulas, scientific principles, the laws of nature, and "mental processes" were to remain excluded from patenting; otherwise, the free pursuit of science would be jeopardized. The "building blocks" of invention were not to be patented. Nevertheless, consensus gradually emerged that inventions of the kind could be eligible for patenting after all—as long as the software was somehow embedded in practical reality. The United States Patent and Trademark Office (USPTO) struggled with several formulations of this requirement over the years (to wit: the algorithm should be applied to physical elements or process steps; the output

[21] Note that EUIPO (2017) also employs the iceberg analogy (on page 14).

should not be "pure number"; does the invention"as a whole" represent more than just building blocks of invention?). In 1996 the case was settled for the first time with the following formula: a demonstrable link with physical reality should exist (USPTO, 1996).

This can be demonstrated to patent application examiners in two ways: either the process requires physical steps that must be performed outside the computer (either before the software runs, or afterwards—the so-called "safe harbours"), or the invention is shown to be tied to "a practical application in the technological arts". Notice that the European Patent Office (EPO) has eventually come round to about the same requirement: a software-related invention—in their terms: a "computer-implemented" invention—should have a "further technical effect" (that is, beyond the technical effect inside any computer; EPO, 2018, G-II-3.6).

As a result, such patenting gradually came to be accepted in the US, and their numbers steadily rose from about the year 2000 onwards. Around 2015 there were several court cases in which the "demonstrable link with physical reality" was put to the test. As a result, many already granted patents were invalidated, some of which were reinstalled later on. In spite of all turmoil, the USPTO continued to examine applications along the lines sketched above.

However, what about inventions related specifically to *ML* (or, broader, AI)? Are they eligible for patenting? After prolonged discussion in patent circles the USPTO in 2019 proposed new guidelines for the thorny problem of patenting abstract ideas—under which heading ML and AI reside—that modernized the former approach to software-related inventions. In these revised guidelines (USPTO, 2019) the class of "abstract idea exceptions" is defined as mathematical concepts, certain methods of organizing human activity, and mental processes (p. 52). If an exception is found to be present in a patent claim (step 2A, part 1; p. 53), examiners should investigate to what extent it relates to a practical application (step 2A, part 2; p. 53). If a claim integrates the exception into a practical application, it is eligible subject matter. By way of example: "an immunization step that integrates an abstract idea into a specific process of immunizing that lowers the risk that immunized patients will later develop chronic immune-mediated diseases" (note 25). Such integration may fail: the application is just an indication that the idea should be performed by a computer, without any more specifications (p. 55, note 30); or just links in general terms to the field of use for the abstract idea in question (p. 55, note 32). Granting a patent would "tie up" the abstract idea and prevent others from using it. However, after failing to satisfy step 2A there is yet another possibility to qualify (step 2B). Consider and evaluate the additional elements (other than the "abstract ideas") in the application. If they are unconventional and go significantly beyond these abstract ideas, the claim does qualify as eligible after all.

In Europe, the current approach is, again, similar. ML methods are deemed non-patentable, since they have an abstract, mathematical nature. To become eligible for protection, a technical effect should be demonstrated. An example given involves "the use of a neural network in a heart-monitoring apparatus for the purpose of identifying irregular heartbeats". However, "classifying text documents solely in respect of their textual content" does not qualify for patentability, since it is regarded as having a linguistic, not a technical purpose (both examples from EPO, 2018, G-II-3.3.1).[22]

This invitation to patent ML-related inventions did not fall on deaf ears—on the contrary. While the stream of such patent filings at the USPTO remained steady from 2000 onwards, at a low level of a few hundred per year, after 2010 it really picked up and has for now reached a maximum of 12,087 (number of applications for 2019; see Fig. 1). Note that, as a rule, about half of them will be granted after examination. For Europe the same tendency applies, though the numbers involved are much smaller. From 2000 onwards about 20 applications per year were filed. Then, after 2010 the numbers rose steadily to 1345 in 2019 (Fig. 2).

The main players in the field of ML as a whole can also readily be identified (see Fig. 3). IBM and Microsoft own most active patents related to ML (about 1500 and 1000 respectively), not unrelated to the fact that they were the first companies to actively pursue such patenting. Amazon and Google follow behind, with almost 500 acquired patents each.

This study is not, of course, about ML in general—it is specifically concerned with profiling applications. From reports and an AI-index produced by WIPO it is possible to derive patterns in the relevant field of "predictive analytics". As can be seen from Figs. 4 and 5, the same general tendency as just observed of rising numbers of ML-related patent applications during the last decade also applies to that specific "functional application". To be more precise: the numbers for the US and for Europe exhibit a steady increase from, respectively, 2014 and 2017 onwards.[23] The WIPO statistics also show that the main players in this field are virtually identical to those in ML at large.

Undeniably, though, this field of patents related to ML has some strange features. Many patents seem to cover a particular ML-method *tout court*. Compare the random forest method invented by Tin Kam Ho and patented by

[22] For a summary comparison of the American and the European patenting landscape, see Tarcu (2019).

[23] Figures 4 and 5 might suggest that the increase for predictive analytics started later than for ML in general, but it has to be borne in mind that these figures report on dates of publication, while Figs. 1 and 2 report on dates of application. As a rule, there is a gap of one, two, or even more years between application and publication.

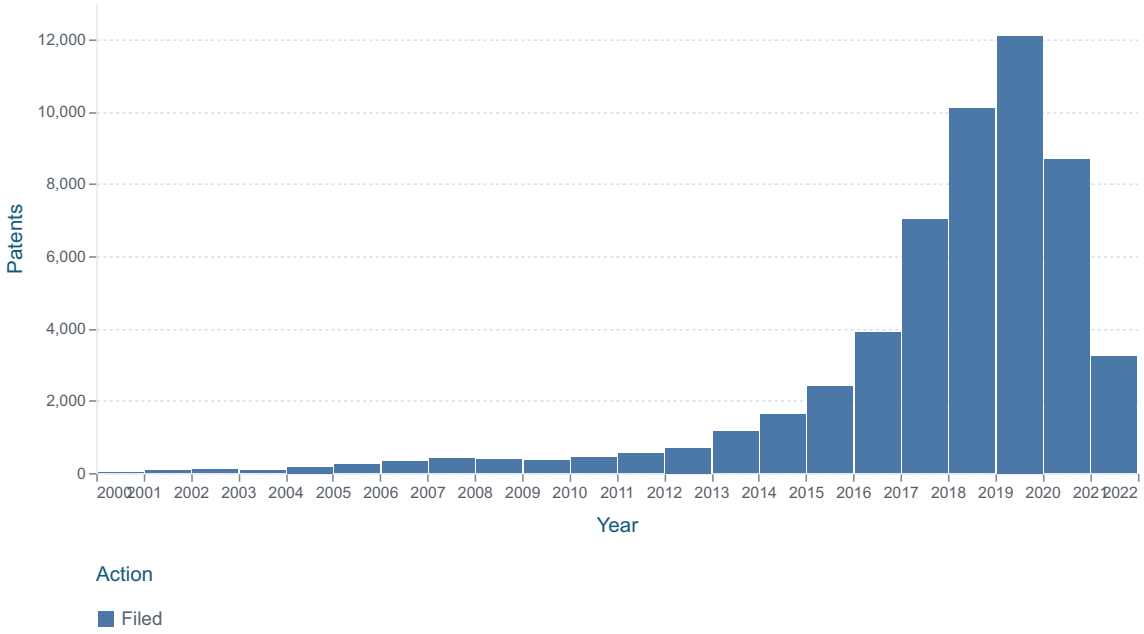

Fig. 1 Annual number of patent applications over the period 2000–2021 with "machine learning" in title, abstract, or claims (US: USPTO). *Source* www.lens.org. Compiled on 14 January 2022.
Notes:
• The search with Lens does not include filings with ML in their *description*, in order not to overestimate the number of machine-learning-related patent applications. If the description is also included in a search—which means that the *complete* text of applications is being searched—the numbers show an approximately fourfold increase.
• The search lists both applications and grants, both active and inactive patents, and has "word stemming" (looking for word variants) turned off.
• The Lens visualisation shown here does not take grouping by patent families into account—which would produce somewhat lower numbers.
• The numbers for 2020 and 2021 do not necessarily indicate a drop in patenting activity, since applications only become public sometime after filing: officially after 18 months. In practice, though, time to publication varies considerably: it may be close to one year (Martin, 2015), but it can also take several years.
• Several other data sources have also been consulted. In particular: the patenting offices themselves (USPTO and EPO) as well as patent data services like AcclaimIP, FPO, and Google Patents. Though the exact numbers may differ, patterns and tendencies over the years are always quite similar

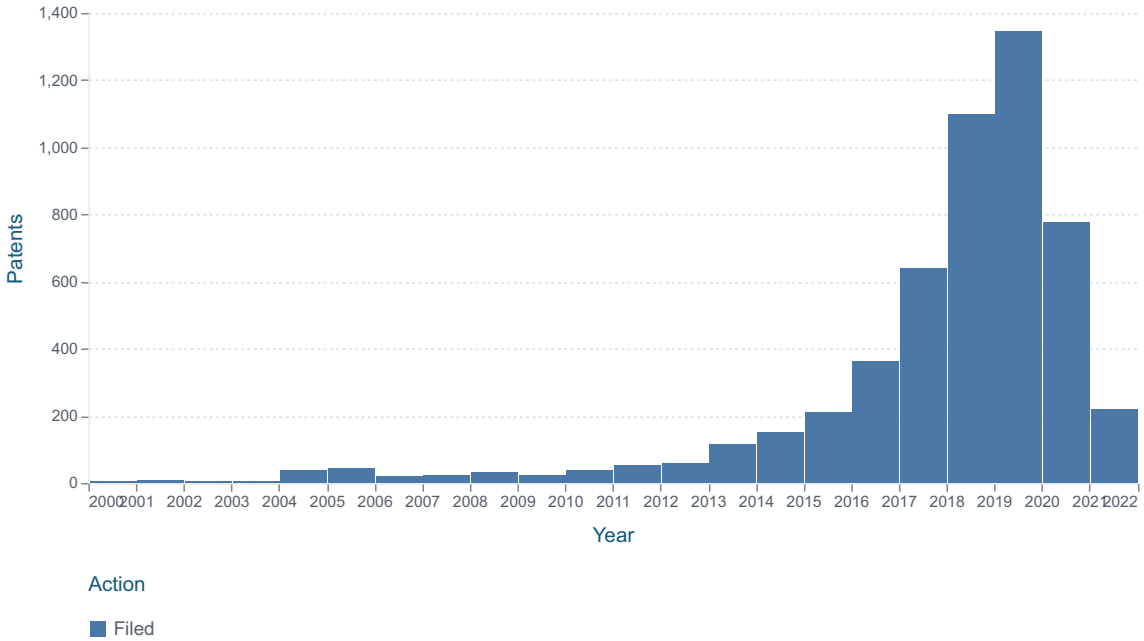

Fig. 2 Annual number of patent applications over the period 2000–2021 with "machine learning" in title, abstract, or claims (Europe: EPO). *Source* www.lens.org. Compiled on 14 January 2022.
Notes: see beneath Fig. 1

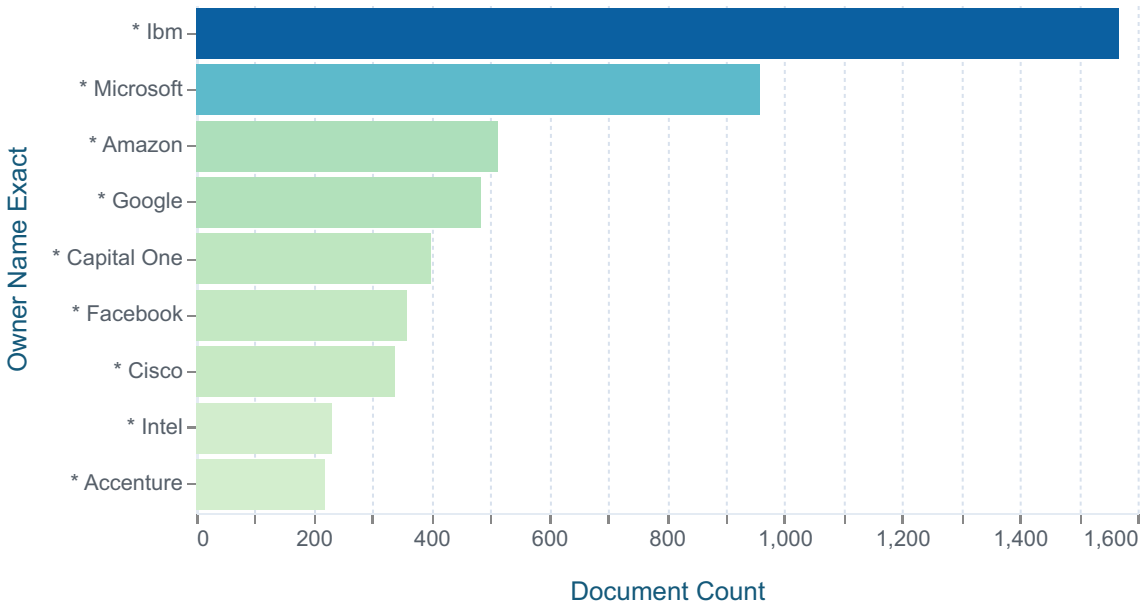

Fig. 3 Companies currently possessing most granted American patents with "machine learning" in title, abstract, or claims (USPTO). *Source* www.lens.org. Compiled on 23 March 2022.
Note: The search with Lens only includes filings (from 1950 onwards) with ML in their title, abstract, or claims. The search lists only granted patents, still active, and has word stemming turned off. Results as shown have not been grouped by families

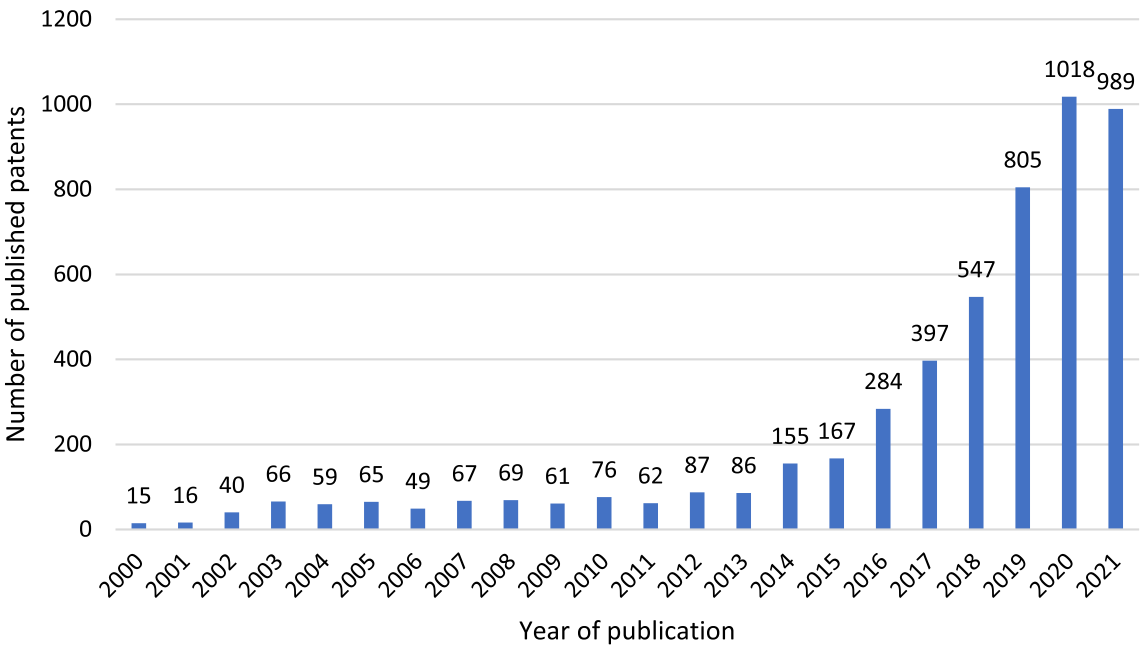

Fig. 4 Annual number of patents published over the period 2000–2021 in the field of "predictive analytics" (US: USPTO, PCT). *Source* WIPO. Compiled on 14 January 2022.
Note: Figure 4 shows patents with application focus on "predictive analytics"; these are almost all based on machine learning. Only one member per patent family is included, word stemming has been turned off. The data are taken from an AI-index containing all patents related to AI ordered into the categories of techniques, functional applications, and application sectors. It has been produced by a team assigned by WIPO and forms the basis for their report about past and present trends in AI (WIPO, 2019). The index is freely accessible and searchable at PATENTSCOPE Artificial Intelligence Index (wipo.int)

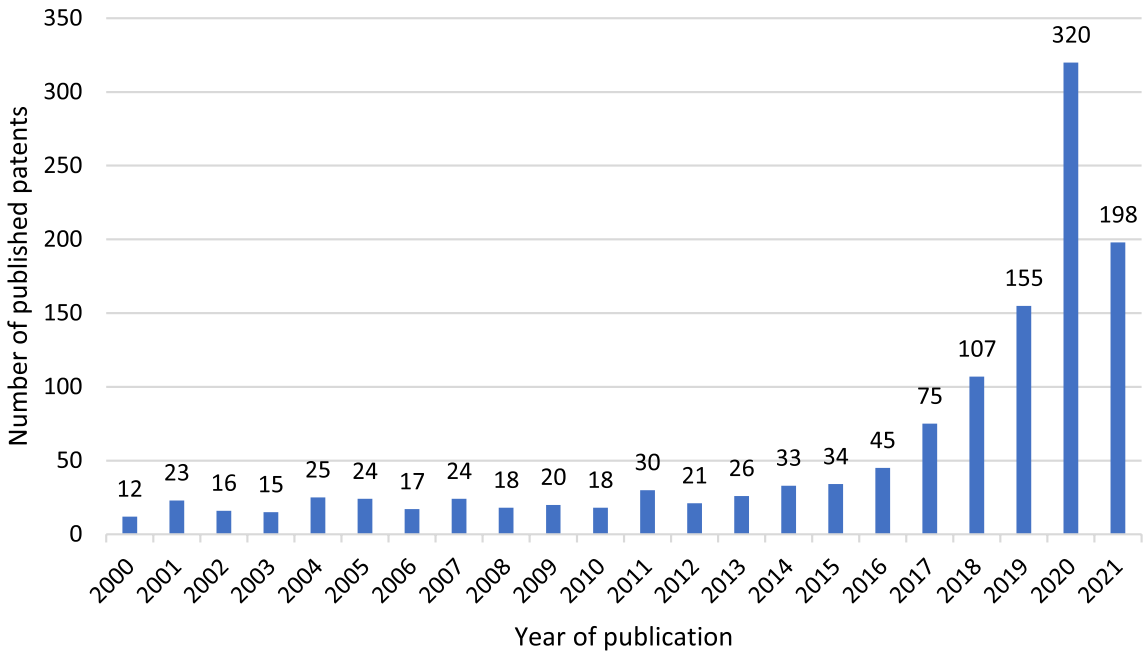

Fig. 5 Annual number of patents published over the period 2000–2021 in the field of "predictive analytics" (Europe: EPO, PCT). *Source* WIPO. Compiled on 14 January 2022. Note: see beneath Fig. 4

Nokia in 1999 ("classification technique using random decision forests", US6009199A). The patent just states that "the invention relates generally to the automatic interpretation of images and patterns, and more particularly to the classification or recognition of such images and patterns." Or the patent related to a support vector machine ("soft margin classifier", US5640492A, invented by Corinna Cortes and Vladimir Vapnik, obtained by Nokia in 1997). The patent publication simply states that it "relates to an apparatus and method for performing two-group classification of input data in automated data processing applications." Just pointing to the field of use conspicuously looks like "tying up" the method per se.

Whenever broad patents of the kind are tested in court, they are almost sure to be invalidated on grounds of ineligibility—their abstract ideas do not seem to be well integrated into a practical application. Many more patents (whether pending or already granted) would meet the same fate, simply because they are found to be obvious, or prior art is brought to the table that destroys the novelty claim. In all, ML-related patents are a large field, but with shaky foundations.

What function is this patenting to serve? In theory, patents serve to protect your turf. If competitors infringe, they are to be sued. In practice though, competitors just proceed with their ML without much regard for each other's patents. Few patent infringement proceedings are actually initiated. So, as the common interpretation goes, firms build large patent portfolios while everybody does so; but this is a defensive measure only—if a patent war ensues, they are prepared to strike back.

Back to the GDPR

In the above I concluded that with the GDPR in force the extent of disclosure about profiling will be determined by a balancing operation: the public interest of data subjects obtaining some knowledge of the algorithmic logic involved as well as the grounds for a decision that pertain to their case against the public interest of honouring the claims for secrecy by data controllers. As we have seen, judicial opinion of European national courts has converged until now on the suitability of *partial* disclosure. After weighing of both interests, a division into two classes of particulars of profiling/scoring was created: vague generalities would have to be disclosed, while specifics of decisions—let alone the software code or the specific model trained—were exempted from disclosure.

As for the present age of the GDPR, I maintain that companies now have *even more* arguments at hand to buttress their secrecy claims. More than ever, the balancing may well tilt in favour of protection of their trade secrets—secret assets that they accumulate in their portfolio of IP assets alongside patents. Organisations like banks or police departments that as a rule source their ICT services from companies, will also advance this argument—as clients they are bound to respect company secrets. In order

to argue this position, I have to go into the details of the required balancing of rights involved.

What kind of weighing is appropriate? What is the proper procedure to be followed in this balancing? In her seminal study *The Foundations of EU Data Protection Law* Lynskey has shown that in its recent judgments the CJEU has offered little guidance on how to balance IP rights with data protection rights (Lynskey, 2015, p. 150 ff.). Invariably, it is stipulated that both kinds of rights are not absolute, their balancing is to be fair and proportionate, and certain limitations on rights may be needed but these should respect the essence of the rights involved. In the end, the CJEU referred the issues brought before them back to the local courts. Instead of this approach that Lynskey qualifies as vague, she suggests that "a preferable approach would be to identify the essential or core objectives of both rights'' (Lynskey, 2015, p. 161). I propose to follow her approach and, while focussing on trade secrecy, interpret its core objective as protection of the competitive advantage that the assets supposedly confer. In other words, the *value* of intellectual property assets in the marketplace is to be protected.[24]

How to conceptualize the value of trade secrets? As far as valuation is concerned, usually three approaches are mentioned (applying to IP assets whether kept secret or patented): *cost* models that refer to the cost incurred for developing the assets, *income* models that try to ascertain the future economic benefits that will accrue from them, and *market* models that aim to determine the market value of the assets involved (cf. EPO, 2011 and Haney, 2020, pp. 471–476). The number of "forward citations" that a patent has generated is often taken as an (indirect) indicator of its income value. As concerns the estimated market value of patents, the well-known survey of 9000 European patents granted in the 1990s (PatVal-EU survey) reported a heavily skewed distribution of values, with a peak around 1 million Euros, 68% less than 1 million Euros, and 32% from 1 million Euros upwards with a long tail (7% was worth more than 10 million Euros) (Giuri et al., 2007, Fig. 4).

For my purposes, only the latter two perspectives are relevant. Upon being forced to open up a trade secret, a company obviously stands to lose much of the asset's income and market value—while the sunk costs to develop the asset just remain sunk. Obviously, though, the costs incurred are an indicator of how much income/market value the firm estimates to reap from the chased invention. For the sake of convenience, henceforth I subsume "income value" and "market value" under the heading "economic value" or simply "value".

Now, as concerns the value of ML assets, I suggest there are two developments that presently contribute to *increasing* the value at stake in such disclosures. For one thing, as ML applications grow in sophistication, companies have to combine several technically related IP assets from their ever-growing pool of ML inventions (either kept secret or patented); they cannot rely on just one invention. In many instances, even assets from *several* companies have to come together. Such "multi-invention contexts" are common (Somaya et al., 2011).[25] Think of neural networks that rely on large scale data sets for training. Several ingenious tricks may be required to execute the training in an efficient and stable way. This observation about sophistication applies in particular to AI services provided by a platform, where customers may upload their data and obtain results ("software-as-a-service", SaaS); that software is always state-of-the-art.

For another, the value of many IP assets is larger than ever since they can be used in several sectors, not just one. This is a characteristic of AI assets (mostly based on ML) in general. As Andrew Ng has remarked: "In the next 5 years, AI will be adopted across multiple industries (especially outside the software industry)" (WIPO, 2019, p. 104). A WIPO report substantiates this claim for patents. Patent applications usually mention a field of application. For predictive analytics in particular, WIPO found that its applications are not only claimed for use in the business sector (as is to be expected), but also in fields such as the life and medical sciences, telecommunications, personal devices (..), and industry and manufacturing (WIPO, 2019, Fig. 3.20). The point to be stressed is that most of those applied patents mention two or more sectors of application.[26] Presumably, the same reasoning about multiple uses is valid for the ML trade secrets in a company's portfolio.

So, with each call by a data subject for an explanation of a specific profiling application, a company will be asked to open up about several IP assets; moreover, each of these assets may well be of use in several other sectors. Then, in court, the firms involved may argue that multiple assets having multiple sectorial uses are at stake, taken together amounting to a considerable value that is entitled to protection. If necessary, the cost incurred for creating the assets may be adduced as an (imperfect) indicator of that (future) value.[27] In those instances, therefore, appeals for trade

[24] Of late, the European Commission, for one, has also emphasized that this is the core of trade secrecy law: cf. https://ec.europa.eu/growth/industry/strategy/intellectual-property/trade-secrets_sv.

[25] This source analyses the complex legal and organisational problems that combination of such distributed inventions has to face. I leave those complications aside.

[26] 62% of all patent applications mention a field of application; 71% of those (=44%) mention two or more fields of application (WIPO, 2019, pp. 49, 51).

[27] As far as organisations (such as bank or police departments) that depend on companies for their profiling activities are asked to open up, they may simply refer the request to those very vendors, with the argument that, as clients, they have to respect their trade secrets.

secrecy protection concerning ML essentials are bound to carry more weight than ever.

If my fears materialize, we are in danger of crossing a critical threshold. In those instances where claims for trade secrecy carry the full weight of multiple assets with multiple uses, we might be approaching the upper limit of exemption: for all practical purposes, the trade secrets exemption from disclosure becomes quasi-absolute. The call for secrecy concerning algorithmic essentials effectively "overrules" the right of access. Disclosures may definitively remain confined to minor details of the profiling systems in use (as has been the practice since 1995). With that, for all practical purposes the public right of access concerning such systems—and *a fortiori* the right to explanation—would be rendered toothless.

Observe that in the GDPR recital that mentions trade secrets, there is also the clause: "However, the result of those considerations should not be a refusal to provide all information to the data subject" (recital 63). This can be read as a warning that trade secrets should not warrant an absolute exemption. But I am not confident this warning will have much practical effect—providing some trivial details will (continue to) suffice.

This is only bound to happen, of course, under one condition. My proposal about what constitutes the essence of trade secrecy and deserves to be considered in any balancing exercise, namely the economic value of the secrets involved, has to gain currency, in one form or another, among corporate lawyers and in legal circles. In particular, in the national courts of Europe and the CJEU. If courts turn out to be insensitive to considerations of the economic value of IP involved, the basis for my fears of trade secrecy protection overriding access rights as a consequence of the steep increase in economic value at stake would evaporate.[28]

A sophist might argue that a climate of heavy intellectual property protection could work just the other way around: towards facilitating disclosures about algorithms. Since database, code, and models are protected by copyright, and database processing and ML processing in use might be covered by patents, one could argue that opening up would hardly prejudice one's competitive position. As the Center for Democracy & Technology suggests on their website: "Companies [like Google] might consider the ability to offer increased transparency" as one of the benefits of forms of intellectual property protection other than trade secrecy (cf. https://cdt.org/issue/privacy-data/digital-decisions/, part VI). However, I think that firms will not readily accept this proposition, for the following reasons. Information disclosures about algorithms or models in use can inspire competitors to create similar algorithms from scratch (circumventing copyright) or "invent around" any patents involved. The most important objection is, though, that essential elements of algorithms in use that do not enjoy protection from copyright or patent law may be considered valuable enough by their creators to be kept as a (trade) secret. Upon disclosure, these elements would no longer be a secret and immediately become available for use by competitors. Valuable competitive advantage gets undermined. I would argue therefore that from the corporate point of view, near-absolute secrecy remains indicated, irrespective of any IPRs obtained.

Critics have suggested to me that the barriers of secrecy might easily be overcome by an "in camera" (in chambers) arrangement: parties to a court dispute open up information, with all present bound to secrecy (an arrangement also mentioned by Wischmeyer, 2019, para 20). Would this constitute an attractive arrangement for the disclosure of algorithmic details, suitable to all parties? I would argue that for the data subjects involved the arrangement is simply too restrictive. At the very least, they want to receive an explanation that enables them to contest a decision if so desired. Moreover, data subjects may have bundled their energies in a class action; such a class is only interested in information that can be used to further their cause. In view of both ends, all information must be available for public use.

In search of evidence

After the GDPR came into force, the Regulation has been implemented at the level of nation states. These incarnations in national law are no ground for much optimism about the right to explanation (as expressed in recital 71) (cf. the extensive analysis in Malgieri, 2019). None of the Member States turns out to have seized the opportunity to mention (in their own language) a right to explanation—only the French law (2018) did so. For fully automated decisions (as far as these are allowed: "administrative" and "private" decisions), besides the usual safeguards an explanation of the individual decision is explicitly due: the data controller should "be able to explain, in detail and in an intelligible form, to the person concerned how the processing has been implemented in his or her individual case" (Article 10 of the French law, translated) (cited in Malgieri, 2019, p. 13).[29]

[28] Recently, another argument against providing explanations for ML outcomes has surfaced. It has been argued that these are food for *adversarial attacks* that target data streams or models. As a result, outcomes of models can be manipulated, or the models themselves stolen (cf. Hind, 2019). This argument would constitute one more reason for companies to keep their ML assets a secret. I leave this argument aside for lack of space.

[29] The Hungarian law states that concerning fully automated decisions the data controller "informs the subject, upon his/her request, of the methods and criteria used in the decision-making mechanism"

Is any evidence available for my conjecture that the exemption for trade secrets is becoming more pronounced and that the right of access and especially to explanation suffer accordingly? In order to answer that question, we would have to turn to case law and analyse developments since 2018 at the level of European states and/or of Europe. Parties referring to the GDPR and its particular national implementation may have brought forward cases concerning the right of access and/or to explanation. I gladly leave that case analysis, being a vast undertaking of its own especially in view of all the languages involved, for other scholars to pursue. After all, my main intention behind this research has only been to alert the reader to the *conjecture* that trade secrecy rights may seriously obstruct the drive towards transparency of algorithmic profiling.

Further, cases of the kind may have landed at the CJEU. This level is of course the most important for putting my conjecture to the test since the Court constitutes the ultimate arbiter for adjudicating court cases of European origin. A preliminary analysis focussing on the CJEU—which was feasible enough—reveals that until now, the right of access, let alone the right to explanation concerning profiling as formulated in the GDPR, have simply not been addressed by the Court.[30] Two reasons suggest themselves. For one thing, four years is simply too short a period for any relevant court case to have been filed. For another, due to prevailing legal opinions from the past about the right of access as described above, reinforced by the national versions of the GDPR (almost) all failing to mention the concept of explanation, no parties have found it worthwhile to take their chances.

Recent evidence about the continuing importance of trade secrecy for companies *is* available, though, albeit not from the European, but from the American continent.[31] Two landmark cases involving profiling and scoring have recently been adjudicated in US courts. The Houston Federation of Teachers asked for suspension of the use of EVAAS, an algorithm for calculating schoolteacher effectiveness scores that influence decisions about bonuses and termination of contract. The other case involved Loomis, a prisoner who asked for algorithmic details of his COMPAS score—COMPAS being a profiling tool used for assessing the chances of recidivism of inmates. In both cases, the companies that created the algorithmic tools, SAS and Northpointe respectively, refused to provide any more details beyond what was already known, arguing that they consider both source code and algorithms to be their intellectual property and wanted to keep them a secret. Financial interests were emphasized; concerning one of the companies (SAS) there were even allusions to possible bankruptcy should they be forced to open up. In both cases, for various reasons, the companies' stance about trade secrecy prevailed.[32]

Conclusion

Regulation by the state as currently practised in Europe is not likely to succeed in venturing far beyond the right of access and granting a proper right to explanation. At least in the short term, citizens subjected to momentous scoring or profiling algorithms may continue to remain in the dark as to the very reasons for outcomes affecting their destinies. This conclusion is based upon the following arguments. As demonstrated by Wachter and co-workers, in the times before the GDPR the trend had been to let trade secrecy considerations take precedence over the right of access. This trend is likely to be strengthened with the GDPR in force, for several reasons. First, the wording of articles and recitals in the Regulation as concerns balancing the right of access against trade secrecy protection is almost identical with the one in the preceding DPD. Moreover, trade secrecy has evolved into a more serious factor to be reckoned with. Its scope has been clarified and its importance emphasized by a fresh directive (EU 2016/943)—algorithms may definitely count as trade secrets. Further, as I have shown, its weight for many a company involved has been increasing in the last decade, since their ML-related applications may combine several secret assets, all of them, moreover, usable in many sectors. If the approach of defining the core objective of trade secrecy as preservation of economic value gains currency in legal circles, the exemption for disclosing trade secrets arguably could approach quasi-absolute status. Thereby the argument from trade secrecy would transform from a protection against competitors into a much broader shield against public scrutiny per se. Trade secrecy would

Footnote 29 (continued)

(section 6). In contrast to Malgieri, this clause does not strike me as providing much of an explanation of an individual decision.

[30] On the Curia website (curia.europa.eu) a search has been performed for relevant cases of European case law. In a series of consecutive searches, I looked for cases referring to GDPR AND (profiling OR algorithm OR explanation OR "right of access" OR "data protection") AND/OR ("trade secrecy" OR "trade secrets" OR copyrights/patents) over the last 5 years (search performed on 28 March 2022). No useful results have been obtained. 40 cases *did* refer to the GDPR, but none of those revolved around the right of access/explanation as concerns profiling, or trade secrets/copyrights/patents—let alone their combination.

[31] The following observations do not imply the suggestion that the respective juridical systems are easily comparable, but only serve to disclose the attitude of the firms involved towards trade secrecy.

[32] The references for the two cases, too numerous to be mentioned here, are available on request.

turn into the perfect excuse for keeping profiling algorithms black-boxed.[33]

Despite these gloomy conclusions, forces *other* than regulation—mainly operating *outside* Europe—can be speculated to be at work that may contribute to effectively providing more insights into algorithmic decisions. For one thing, explainable or (even) interpretable algorithms are urgently demanded in the medical sector. Doctors and specialists continue to feel uneasy about diagnostic systems that are complete black boxes. They argue that they cannot relay a diagnosis to patients without being able to give at least some explication. Further, they are as interested in understanding as in accuracy; they want to gain insights from a learned model, to be used in further development of theory. Also note that accuracy without insight can be misleading: a black box model may have incorporated false correlations from a set of training data that are not immediately evident. No wonder that one overview after another of the application of ML in medicine observes that the black box nature of its algorithms remains a serious obstacle (Shahid et al., 2019; Xiao et al., 2018). While these shortcomings are being noticed, shedding more light on the inner dynamics of, for example, neural networks for clinical applications is making progress (Zhang et al., 2018). Given the considerable funding in this field, progression towards ML methods for generating algorithms with insights included—and introducing these methods in practice—is not unlikely.

For another, in a field some distance away from our discussion of algorithmic decision-making—or medicine for that matter—efforts towards explainable algorithms are building up. It is the American Department of Defence that has concluded it needs such algorithms for modern warfare. The future battlefield will not only enlist soldiers but also robots. If human soldiers are to operate adequately with their robotic companions, as an integrated battle unit embedded in a range of AI systems, they must be able to communicate with these robots and understand them (and vice versa). Similarly, analysts must work together with AI systems that collect and analyse data from the battlefield—think of airplanes and drones.

Precisely for those reasons, DARPA has introduced the Explainable Artificial Intelligence (XAI) program at the end of 2016, which funds research units around the globe with millions of dollars for a period of 4 years. Says the announcement: "The goal of XAI is to create a suite of new or modified ML techniques that produce explainable models that, when combined with effective explanation techniques, enable end users to understand, appropriately trust, and effectively manage the emerging generation of AI systems" (DARPA, 2016, p. 5). Interestingly, their criteria for what counts as an explanation are ambitious. Apart from the model's particular decisions, the strengths and weaknesses of the overall model are also to be explained. Further, the model is to enable the user to identify and correct mistakes.[34]

These demands from the market might generate considerable advances towards explainable ML being developed and used in practice. This speculation can be solidified somewhat by a cursory risk analysis. In the case of the profiling applications for algorithmic decision-making we have been focussing on, decision subjects are affected, but it is usually not a matter of life and death. But in both medicine and warfare the risks *are* considerable: lives are literally at stake. Moreover, powerful actors (doctors and generals) take it upon themselves to push for solutions that reduce those risks—they do not want to jeopardize their patients and their "human resources" respectively. At the end of the day, practical advances in these two sectors that turn explainable ML (or AI) into a feasible option may put pressure on other sectors to follow suit. Companies and governmental organisations in particular that have been using scoring and profiling algorithms for decades—with hardly any explaining taking place—may then finally transform their practices and begin to offer explanations to their decision subjects.

Can, by any chance, powerful market forces achieve what state regulation cannot?

[33] This sombre conclusion about the right to explanation for European citizens can also be of wider interest for other countries and continents, since the GDPR is widely interpreted as setting a standard for other jurisdictions as well (the "Brussels effect").

[34] The foregoing passages about the military need for XAI are not to be read as condoning such applications. These are fraught with all kinds of moral and legal problems that, for reasons of space, have to be ignored here.

References

All websites mentioned below were last accessed on 14 January 2022.

algo:aware. (2018). Raising awareness on algorithms. Procured by the European Commission's Directorate-General for Communications Networks, Content and Technology. Version 1.0, December 2018. https://AlgoAware-State-of-the-Art-Report.pdf (actuary.eu).

Algorithmic Accountability Act of 2019. (2019). House Resolution 2231. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

AlgorithmWatch. (2019). Automating Society; Taking Stock of Automated Decision-Making in the EU. A report by AlgorithmWatch in cooperation with the Bertelsmann Stiftung, supported by the Open Society Foundations. 1st edition, January 2019. www.algorithmwatch.org/automating-society.

Arrieta, A. B., Rodriguez, N.D., Del Ser, J., Bennetot, A., Tabik, S., González, A. B., García, S., Gil-López, S., Molina, D., Benjamins, V.R., Chatila, R., & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. https://arxiv.org/abs/1910.10045.

Article 29 Data Protection Working Party. (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 17/EN, WP251rev.01. ARTICLE29—Item (europa.eu).

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R.P., Moura, J. M. F., & Eckersley, P. (2019). Explainable machine learning in deployment. ArXiv Preprint. https://arxiv.org/pdf/1909.06342.pdf.

Bill 64. (2021). An Act to modernize legislative provisions as regards the protection of personal information. http://www.assnat.qc.ca/en/travaux-parlementaires/projets-loi/projet-loi-64-42-1.html.

C-11. (2020). An Act to enact the Consumer Privacy Protection Act (..). ('Digital Charter Implementation Act, 2020'). https://www.parl.ca/LegisInfo/en/bill/43-2/C-11.

California Consumer Privacy Act (CCPA). (2018). https://oag.ca.gov/privacy/ccpa.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. *Proceedings KDD, 2015*, 1721–1730. https://doi.org/10.1145/2783258.2788613.

Casey, B. (2018). The next chapter in the GDPR's "Right to explanation" debate and what it means for algorithms in enterprise. EU Law Working Papers No. 29, Stanford-Vienna Transatlantic Technology Law Forum.

Citron, D. K., & Pasquale, F. (2014). The scored society: due process for automated predictions. *Washington Law Review, 89*(1), 1–33.

DARPA. (2016). Broad agency announcement. Explainable Artificial Intelligence (XAI). DARPA-BAA-16-53. https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf.

De Laat, P. B. (2000). Patenting mathematical algorithms: What's the harm? A thought experiment in algebra. *International Review of Law and Economics, 20*(2), 187–204.

De Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology, 31*(4), 525–541.

De Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation? *Philosophy & Technology, 34*(4), 1135–1193.

Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review, 16*(1), 1–65.

EPO. (2011). How do you measure patent value? https://www.epo.org/service-support/faq/searching-patents/valuation.html.

EPO. (2018). Guidelines for examination in the European Patent Office. Amended in 2021. https://www.epo.org/law-practice/legal-texts/guidelines.html.

Erkal, N. (2004). On the interaction between patent policy and trade secret policy. *Australian Economic Review, 37*(4), 427–435.

EU. (2020). White paper on Artificial Intelligence: A European approach to excellence and trust. European Commission (europa.eu).

EU. (2021). Proposal for a Regulation COM/2021/206 final (..) laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (..). EUR-Lex - 52021PC0206-EN-EUR-Lex (europa.eu).

EU Directive 95/46/EC (..) on the protection of individuals with regard to the processing of personal data and on the free movement of such data ("Data Protection Directive"). http://data.europa.eu/eli/dir/1995/46/oj.

EU Directive 2016/943 (..) on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure ("Trade Secrets Directive"). http://data.europa.eu/eli/dir/2016/943/oj.

EU Directive 2019/790 (..) on copyright and related rights in the Digital Single Market (..) ("Copyright Directive"). http://data.europa.eu/eli/dir/2019/790/oj.

EU High-Level Expert Group on Artificial Intelligence. (2018–2019). Ethics guidelines for trustworthy AI (draft version in 2018, final version in 2019). Building trust in human-centric AI|FUTURIUM|European Commission (europa.eu).

EU Regulation 2016/679 (..) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (..) (General Data Protection Regulation, GDPR). http://data.europa.eu/eli/reg/2016/679/oj.

EUIPO. (2017). Protecting innovation through trade secrets and patents: determinants for European Union firms. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/Trade%20Secrets%20Report_en.pdf.

Giuri, P., et al. (2007). Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research Policy, 36*, 1107–1127.

Gunst, H. (2017). *The right to explanation and the right to secrecy: reconciling data protection and trade secret rights in automated decision-making*. Master Thesis, University of Helsinki, Finland.

Haney, B. (2020). AI patents: A data driven approach. *Chicago-Kent Journal of Intellectual Property*, 19(3): article 6. https://scholarship.kentlaw.iit.edu/ckjip/vol19/iss3/6.

Hind, M. (2019). Explaining explainable AI. *XRDS, 25*(3), 16–19.

House Resolution 153. (2019). Supporting the development of guidelines for ethical development of artificial intelligence. https://www.congress.gov/bill/116th-congress/house-resolution/153/text.

IEEE. (2018). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (A/IS). Version 2. ead_v2.pdf (ieee.org).

Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal, 34*, 189–218. https://scholar.law.colorado.edu/articles/1227.

Letham, B., Rudin, C., McCormick, T. H., et al. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics, 9*(3), 1350–1371.

Lipton, Z. C. (2016). The mythos of model interpretability. ArXiv preprint. https://arxiv.org/abs/1606.03490.

Lou, Y., Caruana, R., & Gehrke, J. (2012). Intelligible models for classification and regression. *Proceedings KDD, 2012*, 150–158. https://doi.org/10.1145/2339530.2339556.

Lynskey, O. (2015). *The foundations of EU data protection law*. Oxford University Press.

Malgieri, G. (2016). Trade Secrets v Personal Data: A possible solution for balancing rights. *International Data Privacy Law, 6*(2), 102–116.

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Computer Law & Security Law, 35*, 105327.

Malgieri, G., & Comandé, G. (2017). Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law, 7*(4), 243–265.

Martin, J. F. (2015). The myth of the 18-month delay in publishing patent applications. *IPWatchdog*, August 3, 2015. https://www.ipwatchdog.com/2015/08/03/the-myth-of-the-18-month-delay-in-publishing-patent-applications/id=60185.

McKinsey. (2019). Global AI survey: AI proves its worth, but few scale impact. November 2019. https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact.

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining Explanations in AI. [1811.01439] Explaining explanations in AI (arxiv.org).

Molnar, C. (2021). *Interpretable machine learning; A guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/.

New York City Automated Decision Systems Task Force. (2019). Report 2019. https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.

OPC (Office of the Privacy Commissioner of Canada). (2020) A regulatory framework for AI: Recommendations for PIPEDA reform. https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/.

Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

Poursoltani, M. (2020). Disclosing AI inventions. *Texas Intellectual Property Law Journal, 29*, 41–65.

Quinn Emanuel. (2020). The increasing importance of Trade Secret Protection for Artificial Intelligence. https://www.jdsupra.com/legalnews/april-2020-the-increasing-importance-of-64465/.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings KDD, 2016*, 1135–1144. https://doi.org/10.1145/2939672.2939778.

Rudin, C. (2018). Please stop explaining Black Box models for high stakes decisions. https://arxiv.org/abs/1811.10154.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law, 7*(4), 233–247.

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS ONE, 14*(2), e0212356.

Somaya, D., Teece, D., & Wakeman, S. (2011). Innovation in multi-invention contexts: Mapping solutions to technological and intellectual property complexity. *California Management Review, 53*(4), 47–97.

Tarcu, R. (2019). How the EPO and USPTO guidance will help shape the examination of artificial intelligence inventions. IPWatchdog.com, April 1, 2019. http://www.ipwatchdog.com/2019/04/01/epo-uspto-guidance-will-help-shape-examination-artificial-intelligence-inventions/id=107855/.

USPTO. (1996). Examination guidelines for computer-related inventions. Final Version. Federal Register, 61(40), February 8, 1996.

USPTO. (2019). 2019 Revised patent subject matter eligibility guidance. *Federal Register, 84*(4), January 7, 2019. https://www.govinfo.gov/content/pkg/FR-2019-01-07/pdf/2018-28282.pdf. (Last updated in 2020: Manual of Patent Examining Procedure (MPEP), Ninth Edition, Revision 10.2019: par. 2106 Patent Subject Matter Eligibility. https://www.uspto.gov/web/offices/pac/mpep/s2106.html.)

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review, 2019*(2), 494–620.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law, 7*(2), 76–99.

Wachter, S., Mittelstadt, M., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887.

Willoughby, K. W. (2013). Intellectual property management and technological entrepreneurship. *International Journal of Innovation and Technology Management, 10*(6), 1–42.

WIPO. (2019). *WIPO technology trends 2019: Artificial intelligence*. World Intellectual Property Organization.

Wischmeyer, T. (2019). Artificial intelligence and transparency: Opening the black box. In T. Wischmeyer & T. Rademacher (Eds.), *(2019) Regulating artificial intelligence* (pp. 75–101). Springer.

Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association, 25*(10), 1419–1428.

Zhang, Z., Beck, M.W., Winkler, D.A., Huang, B., Sibanda, W., Goyal, H., written on behalf of AME Big-Data Clinical Trial Collaborative Group. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine, 6*(11), 216 ff. https://doi.org/10.21037/atm.2018.05.32.