Explanations in Risk Analysis: Responsibility, Trust and the Precautionary Principle

Salvatore Sapienza[1]

CIRSFID - Alma AI, University of Bologna, Italy

Abstract. The ongoing deployment of machine learning models for public health risk analysis leads to the emergence of concerns regarding the nature of algorithmic-supported decisions with high impact on society or specific groups, with peculiarities that differ from the domain of individual automated-decision making. Such concerns regard the possibility of reinforcing distrust in the institutions responsible for risk assessment and risk management when the nature of their outcomes is not perceived by the society as reliable or trustworthy. This paper proposes a trust-oriented aptitude to the issues emerging from these novel computational practices by discussing the relationship between explanations and the precautionary principle, which governs decision-making processes in the context of public health, and how such relationship can be enhanced by explanations throughout all the steps of risk analysis. The European food safety regulatory framework is taken into account as a significant case study.

Keywords: Explicability · Risk Assessment · Precautionary Principle · Trust · Public Health

1 Introduction

Explainability is an ethical principle primarily intended to promote algorithmic transparency and prevent opaqueness in "Artificial Intelligence" (AI) systems, including machine learning approaches. "Black box" algorithms [25] raise ethical concerns due to the possibility of relying on "inscrutable evidence" when their users take decisions [22]. The potential of machine learning algorithms to support decisions that have an impact on fundamental rights and freedoms has broaden the discussion towards explainability to include the necessity of promoting trust in AI systems, contesting AI-supported decisions, and verifying compliance with fundamental rights [24].

Explainability is also understood as a technical challenge that tries to tackle the opacity of such algorithms bearing effects of automated decisions on individuals [29]. On the one hand, a considerable amount of research - including the eXplainable AI (XAI), research trend - is devoted to ensure that available machine learning techniques guarantee a sufficient understanding of their internal structure (*global* explanation); on the other hand, some have argued that users shall be able to interact with the system to grasp the "why and how" the

decision has been taken (*local* explanation), including by means of intelligible user interfaces [27].

As a principle, Explainability has been endorsed by the European Commission as a key technical requirement [10] functional to the evaluation of fairness [11] of AI systems. The German approach [3] has identified its proximity to trust, as explainable machine learning models allows the assessment of legal compliance [8, 21]. The corollary principle of Explicability has been proposed by AI4People group [1] and the European Commission High Level Expert Group on AI [14], which has referred to this principle as the capability of AI systems to communicate their operations and provide for a rationale for their output.

Legal implications of explainability have been mainly framed within the realm of automated decision-making (ADM) systems that impact on individuals [31]. The discussion has mainly revolved Article 22 of the General Data Protection Regulation (GDPR) (e.g., in [27]) and the contested existence of a Right to Explanation [30]. While focusing on an individual dimension, decisions that do not interfere with specific persons have been partly overlooked by the literature on explainability, perhaps due to the lack of an influential piece of legislation in force that mandates forms of explanations other than the ones used in ADMs. The proposed AI Act[1] explores avenues for the transparency of AI systems which pose high risks for individuals fundamental right (e.g., Recital 38; articles 11, 12, 13) or selected categories of systems (e.g., article 52 and 'deep fakes'). Beside the ongoing legislative procedure, the relevance of a broader perspective can be grounded on the same pillars on which the discussion on individual decision-making systems is currently taking place. In particular, the aforementioned dimensions of trust in machine learning systems, accountability for their use and legal compliance of the decision-makers apply also to the collective and societal implications of inexplicable AI-supported decision-making processes.

This short paper seeks to provide insights on decision-making processes that, although neither interfering directly with individual fundamental rights, nor falling within high-risk categories for the purposes of the AI Act proposal, pose questions with regards to trust, accountability and liability for the use of machine learning systems. Public health risk evaluation in the European Union (EU) represents a challenging research scenario for investigating these social dimensions. Risk-related decisions have a significant impact on the EU population and a certain degree of opaqueness of the underlying scientific risk assessment is perceived by the population. In the case of food safety, this was reported by the 2018 Fitness Check on the General Food Law Regulation [6, §4.2.24]. As a result, a low level of trust in the authorities responsible for food safety risk analysis has been found by the surveys carried out in the drafting of the Fitness Check. Inexplicable machine learning systems can originate an additional layer of distrust, with a negative outcome on the institutional reputation of such authorities. In this paper, a trust-oriented integration of the principles already established by the risk evaluation system is proposed and discussed, with focus

---

[1] Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) Procedure 2021/0106/COD. COM (2021) 206

on the precautionary principle, one of the cornerstones of the EU approach to risk.

Section 2 presents the research framework at stake, with particular regard to the European Food Safety Authority (EFSA) and the food safety system. Then, Section 3 discusses why explanations might be socially desirable also in decision-making processes with relevant collective (i.e., supra-individual) implications, such as risk analysis activities, whereas Section 4 identifies the significant relationship between the precautionary principle and explicability. A short Conclusion summarises the main findings of this paper, discusses its limitations and explores avenues for further research.

2   Machine Learning approaches to Risk Assessment

Risk analysis consists of three steps: a) risk *assessment* concerns the discovery and the identification of unknown possible risks through scientific analysis; b) risk *management* consists of taking informed and scientifically-sound measures that involve a reasonable amount of risk (e.g., authorising or banning the placement of goods in a given market); and c), risk *communication* deals with the proper dissemination of scientific results and disclosure of the rationales underlying the decisions. For instance, in the regulatory domain of European food safety[2], EFSA acts as the risk assessor and it is tasked to provide scientific opinions to the EU Commission, that is the risk manager which takes decisions and measures regarding food-related hazards. Both share risk communication duties. Similar mechanisms are present in medicine and chemicals regulations [5].

Machine learning algorithms for risk classification and prediction are becoming increasingly popular among institutions responsible for risk assessment in the area of public health. Let us consider the study commissioned by the EFSA to explore the potential of this computational approach, including deep neural networks, in its areas of competence [16] and the position paper released by EFSA on its plans regarding data and algorithms [4], and the Memorandum of Understanding between EFSA and the European Chemicals Agency (ECHA) specifically mentioning machine learning [9]. Taken together, these documents highlight a progressive shift from deterministic to probabilistic (or stochastic) forms of risk assessment which poses questions as regards the peculiar nature of their outputs.

In comparison to traditional deterministic methods, scholars [22] have referred to the epistemic concerns of probabilistic algorithms related to inconclusive, inscrutable or misguided evidence used to support decision-making processes. Inconclusive evidence refers to the possible epistemic irrelevance of correlations from a causal perspective; inscrutable evidence broadly refers to the set of issues regarding the data being used and the accessibility of the logic underlying the conclusion reached by the algorithm; misguided evidence refers to the possibility of data-driven fallacies (e.g., gerrymandering or sampling bias)

---

[2] Regulation (EC) No 178/2002 [2002] OJ L 31/1

that might lead to imprecise or incorrect decisions. Public health risk analysis presents similar challenges when machine learning algorithms are deployed by risk assessors to support managers' decision-making.

In the scenario at stake, risk managers cannot directly have an influence over individuals. However, their decisions on how to regulate products can have a significant impact on the society as a whole, in particular as regards long-term effects of regulated substances and products. The same can be said for specific groups which can be more affected than others due to the product's intended use. For instance, in the food safety system, decisions may influence on people following specific diets for religious and philosophical reasons - e.g., vegetarians, vegans or Kashrut - or due to health concerns, e.g., coeliac disease or diabetes. Therefore, Article 22 of the GDPR is inapplicable as no *individual* decision-making is performed, let alone by using personal data or solely by automated means.

Liability gaps represent an additional challenge. Risk assessors are tasked with scientific evaluations and the Court of Justice of the European Union has confirmed the non-binding nature of EFSA scientific opinions in multiple occasions[3] and only "manifest error, abuse of powers or clear excess in the bounds of discretion" can fall within the scrutiny of the Court [2, 13]. When risk assessors act within their scientific mandate, they cannot be held responsible for the decisions taken by risk managers.

Additional layers of complexity are those due to the peculiar nature of risk assessment framework: data transparency is hindered by the need of protecting commercially-sensitive information of entities engaged in R&D activities of innovative products (e.g., chemicals) that submit their data to the risk assessors [15]. Additional concerns also come from the mixed nature of the data processed, which encompass both non-personal, experimental information and personal food consumption data [26].

Carrying on with food safety, a low level of trust in data-based risk analysis processes has already been highlighted by the 2018 Fitness Check of the General Food Law [6], in particular due to the lack of transparency in the evaluation of the studies provided by commercial entities. The deployment of machine learning algorithms can also be seen as the insertion of a further step in the risk analysis chain. The consequences of this insertion on trust in the food safety system - as well as in other neighbour domains such as pharmaceuticals or chemicals - can be significant and are explored in the following section.

---

[3] In Case T-311/06 *FMC Chamical and Arysta Lifesciences v. EFSA* [2008] ECLI:EU:T:2008:205, Court's general order noted that "only measures definitively laying down the position of the institution on the conclusion of that procedure are, in principle, measures against which proceedings for annulment may be brought. It follows that preliminary measures or measures of a purely preparatory nature are not measures against which proceedings for annulment may be brought" (para 43). The exact same wording was used in Court's order in Case T-312/06 *FMC Chemical v. EFSA* [2008] ECLI:EU:T:2008:206 para 43 and Court's order Case T-397/06 *Dow Agrosciences v. EFSA* [2008] para 40

3   Delegated Risk Assessment: Ethical Concerns and Explanations

Consumers do not assess the safety of their foods in ways other than some intuitive checks made using their senses and instinct (for instance, unexpected or bad smell sometimes suggests that food is not edible), or by accessing nutritional information via mandatory labels. While this might be sufficient in daily food consumption, with microbiology and scientific risk assessment developing in parallel with nanotechnology, we have experienced some delegation of certain tasks to scientists capable of evaluating the effects of unknown substances whose *naïve* assessment is not possible. Scholars [19] have referred to the category of 'credence goods' to point out that the quality of certain foods (organic, in their example) cannot be ascertained *ex-ante*. This situation seems not dissimilar to regulated substances which pose unknown risks for human consumption or exposure.

Following food crises, including the 'mad cow' disease in the late 90s, European regulators have intervened to include scientific risk assessment within an institutional framework [2]. It provides that independent scientists work in public health agencies and common individuals *delegate* them those risk assessment tasks which require specific knowledge and precise tools to be performed.

Nowadays, scientific data play a fundamental role in the delegated risk assessment. The EU food safety framework has been subject to a recent reform, i.e., the so-called "Transparency Regulation"[4] aiming at promoting transparency in risk assessment by releasing non-confidential versions of data and documents that support the scientific analysis.

Given the amount of information available for analysis - e.g., experimental and field data submitted by private entities, dietary intake information, scientific literature, data from previous assessments, public health statistics (e.g., hospitalisation rates) - risk assessors delegate to machine learning algorithms the tasks that involve large, mutable and heterogeneous data, thus requiring computations capable of "making sense" of them. In this scenario, we confront with a *delegating delegation* situation.

It is then necessary to clarify how such form of delegation to machine learning algorithms differs from the use of support tools. On the one hand, a scientist outsources a portion of a task to the microscope (tool) aiming to make her or his duty (i.e., the observation of a phenomenon) feasible; on the other hand, outsourcing something to a machine learning system entails the delegation of the whole task to a tool capable of minimising the efforts needed to achieve the risk assessor's intended goal. While the microscope *supports* the risk assessor in the fulfilment of a given task, the scientist *delegates* the whole task to a machine learning algorithm. Within AI ethics research, the centrality of trust

---

[4] Regulation (EU) 2019/1381 of the European Parliament and of the Council of 20 June 2019 on the transparency and sustainability of the EU risk assessment in the food chain and amending Regulations (EC) No 178/2002, (EC) No 1829/2003, (EC) No 1831/2003, (EC) No 2065/2003, (EC) No 1935/2004, (EC) No 1331/2008, (EC) No 1107/2009, (EU) 2015/2283 and Directive 2001/18/EC [2019] OJ L231

has been justified by the necessity of discussing issues such as the loss of human control over the execution of the task and the necessity of a human oversight under various models [14]. When contextualised to the risk assessment scenario presented here, the human intervention over the algorithms is limited to the *ex-ante* programming of the code to be executed and the *ex-post* observations of the results.

In both delegation processes - i.e., the one from individuals to risk assessors and the one from risk assessors to machine learning system - trust is deemed to be a key element also due to the nature of the food as credence good [19] and the constant trend that occurs every time algorithms are proven to be unreliable [12]. Trust in the algorithm is a crucial component of the aforementioned kind of delegation [28]. From a governance perspective, the idea of "AI Trustworthiness" has been one of the key drivers of the ongoing regulatory process in the EU [10]. In correlation to the loss of human oversight, overestimating the potentialities of machine learning might generate excessive expectations towards probabilistic results generated with little or no scrutiny over the logic followed by the algorithm used to perform a delegated task [28, 22].

This trust-oriented approach cannot be merely seen as the key driver of the final step. If the evaluation conceptualised as an unified decision-making process from a governance perspective, each step of risk analysis shall be functional to enhcance trust in the system. Trust can serve as a middle-out principle [23] that governs every step of risk evaluation - including the deployment of machine learning algorithms - from the scientific assessment to the perception of the decision by the general public. Explanations represent a key element of such trust-based approach as they are functional to take informed decisions that can be perceived as sound, robust and reliable by the public.

Explainability is desirable in the scientific and institutional scenario at stake. This seems evident when considering that the lack of trust already shown by the 2018 Fitness Check might be combined with the general scepticism towards outputs generated by machine learning algorithms that are perceived as inexplicable by the scientific community. Through its voice, distrust might be spread in the public opinion, ultimately bringing negative effects on the whole sector, including both the Authority and the market. EFSA's commissioned study [16] seems aware of the issues regarding algorithmic transparency as it reports "transparency scores" for each algorithm, even without providing details on the scoring metrics.

Striking a balance between the responsiveness and the certainty of the analysis is also a massive challenge due two clashing interests at stake. While the industry calls for faster "smart" approval procedures, certainty of the results cannot be achieved without cautious scrutiny and extensive cross-validation tests, which naturally require time to be performed. The precautionary principle, discussed in the next section, has been traditionally considered as the guideline to strike this balance. Hoverer, the introduction of novel analytics techniques - including machine learning - calls for an adapted interpretation of the principle.

4   The Precautionary and the Explicability Principles

The precautionary principle has been defined as the necessity that, "in cases of serious or irreversible threats to the health of humans or ecosystems, acknowledged scientific uncertainty should not be used as a reason to postpone preventive measures" [17, 20]. In data-driven risk assessment, the principle can be understood as the case in which the lack of data originates scientific uncertainty towards a potential threat. When this scenario occurs, preventive measures have to be taken by risk managers. *Inter alia*, the European Court of Justice in *Alpharma*[5] framed the relationship between risk assessment and risk management in the light of the precautionary principle when the evaluation revolves around data and forecasts. The Court observed that "[N]otwithstanding the existing scientific uncertainty, the scientific risk assessment must enable the competent public authority to ascertain, on the basis of the best available scientific data and the most recent results of international research, whether matters have gone beyond the level of risk that it deems acceptable for society" (para 175).

Explicability is an ethical principle related to explainability [18] first proposed by the AI4People initiative [1] and it is composed by the sub-principles of "intelligibility" of the AI systems and "accountability" for their use. It expresses the practice to select an appropriate level of abstraction that fulfils the desired explanatory purpose, is appropriate to the system and the receivers of the explanation, deploys suitable persuasive arguments and finally provides explanations to the receiver of its outputs [7]. Explicability thus becomes necessary every time AI outputs need to be understood and interpreted by their receiver. In the scenario at stake, three levels of receivers can be found: first, the risk assessor qualifies as receiver when reading the results of the analysis (*evidence* level); then, risk managers receive the evidence-supported scientific report by the risk assessor and develop their opinion on this basis (*information* level); finally, individuals observe the final output of the analysis, i.e., the concrete measure taken by the risk manager (*decision* level).

A progressive shift from a focus on intelligibility of the systems to the accountability for their use can be also observed: while the risk assessors qualifies as a technical receiver that should be able to scrutinise the output of the algorithm, the general public is more concerned with the accountability of the decision-makers. Risk managers lie somewhere in-between because they are both accountable for their decisions and interested in understanding the nature and the scale of the hazard.

At the *evidence* level, a careful consideration of explicability suggests that if the machine learning models deployed in risks assessment activities do not provide sufficient justifications for their results, it follows that risk assessors will confront with a *ignotum per ignotius* situation, i.e., the scenario in which an explanation is more cumbersome than the phenomenon that the one it should

---

[5] Case T-70/99 Alpharma Inc. v Council of the European Union ECLI:EU:T:2002:210 [2002] ECR-II-03495

clarify. This is not desirable in light of the goal of risk assessment, i.e., reducing uncertainty towards certain phenomena.

At the *information* level, the principle of precaution binds risk managers. Then, they can allow for less cautious decisions only if the layers of uncertainty surrounding certain risks are erased to a threshold that is the level of social acceptance of any risk. When unexplainable machine learning models prevent a deep scrutiny of possible risks due to the inscrutability of their working, the precautionary principle still applies. In other words, the opacity of machine learning models shall be considered an integral part of the scientific uncertainty that might lead to the "most cautious decision" to be taken by risk managers in accordance with the precautionary principle. Such conceptualisation of the scientific uncertainty explains the correlation between explainability and precaution. Their relationship can be constructed as follows: the "most cautious decision" has to be taken every time the level of uncertainty is intolerably high due to the use of opaque machine learning algorithms, according to a precautionary evaluation.

This correlation can also be translated into governance models and policy-making. When machine learning algorithms are deployed but their outcome is characterised by an unacceptable levels of opacity, a comparable conventional or deterministic method should be also used to cross-check the validity of the results. Contextualising algorithmic opaqueness in terms of scientific uncertainty - yet, within risk analysis procedures - can also fruitfully ease the attribution of liability by providing a guideline for the cases in which risk managers neglected the results of algorithmic scrutiny. This novel interpretation can still occur within the current legislative framework, as it does not diverge from the existing responsibility schemes but it simply offers a re-conceptualisation of the link between epistemic aspects of decision-making processes and their legal assessment.

Finally, at *decision* level, the narrative towards explicability has usually been constructed as the instrumentality of explanations to the exercise of the right to contest individual automated decision-making by some authors [24, 31]. In the domain under scrutiny, the right to contest the decision is not applicable since the data analysis does not impact on individual data subjects. Nonetheless, explanations can contribute to the social acceptability of decisions taken by risk managers, in line with what has been recommended by the institutional AI charters mentioned above. Social acceptability also requires appropriate presentation and design strategies to be implemented within the framework of risk communication.

5   Final remarks

This short paper has proposed a conceptual framework to interpret explanations in the context of public health risk assessment, with specific regards to food safety. Following a short review of the ethical perspective of explainability and explicability, this study has argued in favour of a desirability of explanations also in the context of public health risk analysis. This claim has been justified by the

need of fostering trust in the whole institutional framework at every step of the evaluation, considering the low level of trustworthiness shown by recent surveys. Moreover, explanations can serve as an additional parameter for the evaluation of scientific uncertainty when the precautionary principle has to be invoked as a regulatory mechanism. This can ultimately bring beneficial consequences to the attribution of accountability on decision-makers.

As a conceptual paper, this study suffers from limitations due to its narrow scope of analysis and its broad approach to explainability/explicability. On the one hand, limiting the scope to food safety implies that the analysed shift towards machine learning techniques is still an ongoing process, whereas other "neighbour" domains can offer more consolidated scenarios. At the same time, food safety is very unique due to the recent legislative scrutiny over data-related questions, hence prone to low generalisability. On the other hand, literature on explainability/explicability is vast and several theories about explanations have been proposed. This paper does not provide details on the right approach to explanations to be followed, but it is limited to identify *why* explanations in the context of risk analysis are desirable from a trust-oriented perspective and in relation to the precautionary principle.

Considering these limitations, "neighbour" domains to food safety (e.g., chemicals, pharmaceuticals) should be evaluated and juxtaposed in order to formalised a unique framework. The *how* explanations should work will be also left to further research, which should identify what theory of explanation works best within the current scenario, modulate the design requirement for an effective communication and set practical guidelines for implementation. Moreover, the ongoing legislative process of the AI Act hinders an easy categorisation of the AI systems at stake and prevents more clarity on the transparency requirements. Therefore, further research could explore compliant implementations at the end of the legislative process of the AI Act.

References

[1] AI4People. *AI4People | Atomium*. 2018. URL: https://www.eismd.eu/ai4people/.

[2] Alberto Alemanno and Simone Gabbi. *Foundations of EU food law and policy: Ten years of the European food safety authority*. Routledge, 2016.

[3] German Federal Ministry for Economic Affairs BMWi and Energy. *Artificial Intelligence Strategy*. https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20180718-key-points-for-federal-government-strategy-on-artificial-intelligence.html. (Accessed on 02/11/2020). 2018.

[4] Stefano Cappè et al. "The future of data in EFSA". In: *EFSA Journal* 17.1 (2019), e17011.

[5] Damian Chalmers, Gareth Davies, and Giorgio Monti. *European union law*. Cambridge university press, 2019.

[6]   EU Commission. *REFIT Evaluation of the General Food law (Regulation (EC) No 178/2002' SWD(2018) 38 final,* `https://ec.europa.eu/food/sites/food/files/gfl_fitc_comm_staff_work_doc_2018_part1_en.pdf`. (Accessed on 05/15/2021). 2018.

[7]   Josh Cowls et al. "Designing AI for social good: Seven essential factors". In: *Available at SSRN 3388669* (2019).

[8]   Finale Doshi-Velez et al. "Accountability of AI under the law: The role of explanation". In: *arXiv preprint arXiv:1711.01134* (2017).

[9]   EFSA - ECHA. *Memorandum of Understanding ECHA - EFSA.* `https://www.efsa.europa.eu/sites/default/files/assets/mouecha.pdf`. (Accessed on 05/14/2021). 2017.

[10] European Commission. *Commission White Paper on Artificial Intelligence - A European approach to excellence and trust.* `https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf`. (Accessed on 05/22/2020). 2020.

[11] European Commission. *Communication Artificial Intelligence for Europe.* `https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe`. (Accessed on 02/11/2020). 2018.

[12] Fabio Fossa. "«I Don't Trust You, You Faker!» On Trust, Reliance, and Artificial Agency". In: *Teoria. Rivista di filosofia* 39.1 (2019), pp. 63–80.

[13] Simone Gabbi. "The European Food Safety Authority: judicial review by community courts". In: *Revue européenne de droit de la consommation (REDC)-European Journal of Consumer Law* 1.2009 (2008), pp. 171–189.

[14] High Level Expert Group on AI HLEG. "Ethics guidelines for trustworthy AI". In: *B-1049 Brussels* (2019).

[15] Martin Holle. "The Protection of Proprietary Data in Novel Foods–How to Make It Work". In: *European Food and Feed Law Review* (2014), pp. 280–284.

[16] IZSTO et al. "Machine Learning Techniques applied in risk assessment related to food safety". In: *EFSA Supporting Publications* 14.7 (2017), 1254E.

[17] Sheila Jasanoff. *The ethics of invention: technology and the human future.* WW Norton & Company, 2016.

[18] Anna Jobin, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines". In: *Nature Machine Intelligence* 1.9 (2019), pp. 389–399.

[19] Robert Lee. "Novel Foods and Risk Assessment in Europe". In: *The Oxford Handbook of Law, Regulation and Technology.* 2017.

[20] Marco Martuzzi, Joel A Tickner, et al. *The precautionary principle: protecting public health, the environment and the future of our children.* 2004.

[21] Brent Mittelstadt, Chris Russell, and Sandra Wachter. "Explaining explanations in AI". In: 2019, pp. 279–288.

[22] Brent Mittelstadt et al. "The ethics of algorithms: Mapping the debate". In: *Big Data & Society* 3.2 (2016), p. 2.

[23] Ugo Pagallo, Pompeu Casanovas, and Robert Madelin. "The middle-out approach: assessing models of legal governance in data protection, artifi-

cial intelligence, and the Web of Data". In: *The Theory and Practice of Legislation* (2019), pp. 1–25.

[24]   Monica Palmirani. "Big Data e conoscenza". In: *Rivista di filosofia del diritto* 9.1 (2020), pp. 73–92.

[25]   Frank Pasquale. *The black box society.* Harvard University Press, 2015.

[26]   Craig Simpson. "Data Protection under Food Law Post: in the Aftermath of the Novel Foods Regulation". In: *Eur. Food & Feed L. Rev.* 11 (2016), p. 309.

[27]   Francesco Sovrano, Fabio Vitali, and Monica Palmirani. "The difference between Explainable and Explaining: requirements and challenges under the GDPR". In: (2019).

[28]   Mariarosaria Taddeo and Luciano Floridi. "How AI can be a force for good". In: *Science* 361.6404 (2018), pp. 751–752.

[29]   Cédric Villani, Yann Bonnet, Bertrand Rondepierre, et al. *For a meaningful artificial intelligence: Towards a French and European strategy.* Conseil national du numérique, 2018.

[30]   Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. "Why a right to explanation of automated decision-making does not exist in the general data protection regulation". In: *International Data Privacy Law* 7.2 (2017), pp. 76–99.

[31]   Sandra Wachter, Brent Mittelstadt, and Chris Russell. "Counterfactual explanations without opening the black box: Automated decisions and the GDPR". In: *Harv. JL & Tech.* 31 (2017), p. 841.