The European Legal Framework for Medical AI

David Schneeberger[1,2(✉)], Karl Stöger[1], and Andreas Holzinger[2]

[1] University of Graz, Universitätsstrasse 15, 8010 Graz, Austria
{david.schneeberger,karl.stoeger}@uni-graz.at
[2] Medical University of Graz, Auenbruggerplatz 2, 8036 Graz, Austria
andreas.holzinger@medunigraz.at

Abstract. In late February 2020, the European Commission published a White Paper on Artificial Intelligence (AI) and an accompanying report on the safety and liability implications of AI, the Internet of Things (IoT) and robotics. In its White Paper, the Commission highlighted the "European Approach" to AI, stressing that "it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection". It also announced its intention to propose EU legislation for "high risk" AI applications in the nearer future which will include the majority of medical AI applications.

Based on this "European Approach" to AI, this paper analyses the current European framework regulating medical AI. Starting with the fundamental rights framework as clear guidelines, subsequently a more in-depth look will be taken at specific areas of law, focusing on data protection, product approval procedures and liability law. This analysis of the current state of law, including its problems and ambiguities regarding AI, is complemented by an outlook at the proposed amendments to product approval procedures and liability law, which, by endorsing a human-centric approach, will fundamentally influence how medical AI and AI in general will be used in Europe in the future.

Keywords: Anti-discrimination · EU legal framework · Explainability · Fundamental rights · GDPR · Human dignity · Human in the loop · Informed consent · Liability · Medical AI · Product approval · Right to explanation

1 Fundamental Rights as Legal Guidelines for Medical AI

1.1 Some Basic Information on Fundamental Rights in the EU

(European) fundamental rights (a.k.a. human rights) constitute part of the highest "layer" of EU legislation ("primary law") and provide already today an important legal (and not merely ethical) basic framework for the development and application of medical AI. Lower layers of EU law ("secondary and tertiary law") have to respect the guidelines of this framework which is – in contrast

Published by Springer Nature Switzerland AG 2020
A. Holzinger et al. (Eds.): CD-MAKE 2020, LNCS 12279, pp. 209–226, 2020.
https://doi.org/10.1007/978-3-030-57321-8_12

to these lower layers – not very likely to change substantially in the coming years. The main source of this framework is the European Charter of Fundamental Rights (CFR), which is in its entirety applicable to the use of medical AI because the provision of medical services is covered by the freedom to provide services under European law. For its part, the CFR is strongly modelled on the European Convention on Human Rights (ECHR), which is also applicable in all EU states. In spite of the diversity of national legislation in the EU Member States, these two instruments ensure a rather uniform level of protection of fundamental rights across the EU. As medical AI can affect a person's physical and mental integrity in a very intense way and any malfunction could have serious consequences, it is a particularly relevant field of AI in terms of fundamental rights.

In this context, it should be stressed that fundamental rights not only protect individuals from state intervention, but also oblige the state to protect certain freedoms from interference by third parties. The state can fulfil these so-called "obligations to protect" by, for example, enacting appropriate legislation that applies to relations between private individuals or by creating specific approval procedures for placing goods or services on the market that could endanger the fundamental rights of its users. This is why "obligations to protect" are of particular importance in medicine: For example, the European Court of Human Rights has repeatedly stated that fundamental rights entail an obligation on the state to regulate the provision of health services in such a way that precautions are taken against serious damage to health due to poorly provided services [33]. On this basis, the state must, for example, oblige providers of health services to implement quality assurance measures and to respect the due "standard of care". To sum up, fundamental rights constitute a binding legal framework for the use of AI in medicine which is not only relevant for EU Member States, but for all developers and providers of medical AI.

1.2 Human Oversight as a Key Criterion

It has already been emphasized by the Ethics Guidelines of the HLEG that "European AI" has to respect human dignity, one of the key guarantees of European fundamental rights, (Art. 1 CFR) [34], which means that medical AI must never regard humans as mere objects [15]. Every human being therefore has a right that the state respects and protects his or her individuality, also towards third persons. Humans must therefore "never be completely or irrevocably subjected to technical systems" [13], which also applies to the use of medical AI. Since AI works on the basis of correlations, complex AI applications in particular must always be monitored by human beings to ensure that they do not miss any special features of human thinking or decision-making. From this, it can be deduced that the demands for human oversight expressed in computer science [20] are also required by EU fundamental rights. Decisions of medical AI require human assessment before any significant action is taken on their basis. The European Union has also implemented this fundamental requirement in the

much-discussed provision of Art. 22 General Data Protection Regulation (henceforth GDPR), which allows "decisions based solely on automated processing" only with considerable restrictions (discussed in further detail in the following Sect. 2.2 about the GDPR). In other words: European medical AI legally requires human oversight (a.k.a. "a human in the loop" [35]).

1.3 Medical AI and Anti-discrimination Law

There is a rich body of fundamental rights provisions requiring equality before the law and nondiscrimination, including gender, children, the elderly and disabled persons in the CFR (Arts. 20–26). From these provisions, further requirements for the development and operation of European medical AI can be deduced: Not only must training data be thoroughly checked for the presence of bias, also the ongoing operation of AI must be constantly monitored for the occurrence of bias. If medical AI is applied to certain groups of the population that were not adequately represented in the training data, the usefulness of the results must be questioned particularly critically [27,29]. At the same time, care must be taken to ensure that useful medical AI can nevertheless be made available to such groups in the best possible way. In other words: European medical AI must be available for everyone. The diversity of people must always be taken into account, either in programming or in application, in order to avoid disadvantages.

1.4 Obligation to Use Medical AI?

However, fundamental rights not only set limits to the use of AI, they can also promote it. If a medical AI application meets the requirements just described, it may also be necessary to use it. European fundamental rights – above all the right to protection of life (Art. 2 CFR) and private life (Art. 7 CFR) – give rise to an obligation on the part of the state, as already mentioned above, to ensure that work in health care facilities is carried out only in accordance with the respective medical due "standard of care" (a.k.a. "state of the art") [57]. This also includes the obligation to prohibit medical treatment methods that can no longer be provided in the required quality without the involvement of AI [53]. This will in the near future probably hold true for the field of medical image processing.

2 Some Reflections on Four Relevant Areas of "Secondary" EU Law

2.1 Introduction

While the European fundamental rights described above constitute a basic legal foundation for the use of medical AI, the details of the relevant legal framework need to be specified by more detailed legislation known as "secondary law" (e.g.

directives and regulations) or by "tertiary law" (which specifies secondary law even further). The main purpose of this legislation is to create a more or less uniform legal framework across all EU Member States either by replacing national legislation (in particular through regulations) or by harmonizing its contents (in particular through directives). Hence, while the picture we have presented so far has been painted with a broad brush, we will now take a more in-depth look at the finer details of three areas of EU "secondary law", which are of specific importance for the use of medical AI: the GDPR, product approval procedures and the question of liability. Some conclusions we have already drawn in the section about fundamental laws (e.g. human oversight as a key criterion) are further strengthened and expanded through this more detailed overview.

2.2 The General Data Protection Regulation (GDPR) and the "Right to an Explanation"

The GDPR establishes transparency as a key principle for data processing and links it with lawfulness and fairness (Art. 5 para 1(a) GDPR) which both are important parts of the principle of accountability (Art. 5 para 2 GDPR). This focus on transparency as a basic requirement for data processing should be kept in mind when discussing the GDPR.

The presentation of the first draft of the GDPR marked the starting point of an extensive "right to explanation" debate in legal academia (e.g. [6,7,9,11,18,19,28,38,42,43,55,58,59]) the implications of which were also felt in computer sciences [40]. To (mostly) circumnavigate this intricate and complicated debate, which is muddled in semantics, we will not enter into the academic discussion about what constitutes or does not constitute an explanation - which is still a point of contention [41,44,46] - but we will try to clarify the duties to provide information without too much speculation.

Prohibition of Decisions Based Solely on Automated Processing. As stated above, the aim of Art. 22 GDPR is to prevent that individuals will be regarded as mere objects in an automated decision-making process determined solely by machines. Such a situation would result in the loss of their autonomy and hence the loss of human control and responsibility [9]. Therefore Art. 22 para 1 GDPR provides for a prohibition of autonomous decision-making without human assessment ("solely based on automated processing"), the final decision should always remain in human hands.

Decision-support systems are not affected by this prohibition, as long as the human in the loop has substantial powers of assessment and can change the outcome (e.g. doctor who decides based on an AI recommendation). However, if the human does not have any real authority to question the outcome (e.g. nurse who is obliged to strictly follow the AI recommendations) this equals to prohibited fully automated decision-making [2].

If an AI-system is designed for a fully autonomous approach, it will only be prohibited if the decision has serious consequences (a legal effect or a similarly

significant effect) [2]. In the context of medical AI for diagnosis or treatment this threshold will almost certainly be reached, therefore medical AI without a human in the loop is generally prohibited under the GDPR regime. However, there are a few exceptions. The most important exception in the context of medical AI is the documented (e.g. written/electronic) explicit consent of the "data subject" (that is the patient) to the fully automated processing of their health data (data related to physical or mental health [Art. 22 para 4 GDPR]) [23]. As the principle of "informed consent" is (also outside the scope of data protection law) one of the pillars of medical law not only in EU law, but also in the law of EU Member States, there is only one further exception to the requirement of "informed consent" which is, however, to be interpreted narrowly: Automated processing of health data may take place in the reasons of substantial public interest, e.g. public health. Under this exception, it would e.g. be conceivable to identify persons that are particularly vulnerable to a pandemic disease like COVID-19 by means of a fully automated AI system. However, it should be stressed again that this exception is only applicable if a substantial public interest shall be protected. Consequently, it must not be used as a blanket exception to easily circumvent the prohibition stated by Art. 22 para 1 GDPR (Recital 71 lists fraud and tax-evasion monitoring and prevention purposes as example cases) [10,31].

Human in the Loop as a Necessary Safeguard. Even if explicit consent was gained, additional safeguards to protect the rights and freedoms of the data subject must be implemented. The GDPR does not state an exhaustive list of such safeguards, it only lists three examples of these, which constitute a bare minimum standard: a. the right to obtain human intervention, b. to express one's point of view and c. to contest the decision. Accordingly, even in cases where fully automated decision-making is in principle permissible, human intervention and human assessment are still required. Furthermore, the processing of health data as a particularly sensitive category of data requires a higher standard and the implementation of additional safeguards (e.g. frequent assessments of the data to rule out bias, to prevent errors etc.) [2]. Therefore, the GDPR necessitates human oversight a.k.a a human in the loop for medical AI, irrespective of whether it is designed as a decision-support or a fully automated system.

Is There a "Right to an Explanation"? The accompanying so-called recitals of the GDPR, which function as interpretative guidelines [43], also mention additional safeguards for the fully automated processing of health data, among them "The right to obtain an explanation of the decision reached after such assessment" (Recital 71). At the first look this clearly seems to be a right to an explanation of an individual decision, but as recitals are primarily interpretative in nature, this "right" is (according to the current reading) more or less a recommendation and not an obligation. The "controller" (the natural or legal person responsible for the processing of data) is free to choose the safeguards it deems necessary as long as three basic safeguards (possibility of human intervention,

expression of the data subject's point of view, contestation of the decision) are upheld, compliance with the GDPR is present [30,58]. While the implementation of a "right to an explanation" ultimately is not obligatory, it is seen as a good practice to foster trust and is recommended in the GDPR guidelines [2].

Consequently the answer to the question "Is there a right to an explanation?" depends on the definition of "explanation". If "explanation" is defined as "information about basic system functionality" (see the following paragraph), the answer is in the affirmative, if explanation is interpreted broadly in the sense of "explain the causes/internal processes which lead to an individual decision", the answer is in the negative. As long as either the European Court of Justice does not clarify and expand the "right to an explanation" (e.g. the "right to be forgotten" was also created primarily through interpretation by the court [26]) or the GDPR is amended, it seems to only be a recommendation (and, as stated above, mere decision-support systems do not fall under the scope of such a right).

Duty to Provide Information/Right to Access. Even if there is no obligation to explain a specific decision, any controller of an AI-system based solely on automated processing nevertheless has, according to Arts. 13 and 14 GDPR, to provide the subject - in addition to basic information - with information about 1. the existence of automated decision-making, 2. meaningful information about the logic involved (comprehensive information about the reasoning/system functionality, e.g. models, features, weights etc.) and 3. about the significance and the envisaged consequences of such processing (e.g. use for detection of melanoma). Even outside the scope of fully automated decision-making, where there is no obligation to provide this information, it should nevertheless be provided voluntarily as a good practice to ensure fairness and transparency. Independently of the duty of the designer/user to provide information, Art. 15 GDPR grants the data subject a symmetrical right of access to the information defined in Arts. 13 and 14 GDPR (logic involved, significance and envisaged consequences) [2]. Art. 12 GDPR clarifies that this information must be provided in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child (if it is already required to give "informed consent"). The information should normally be provided in a written form (including electronic means), free of charge and without undue delay (maximum within one month) [3]. Therefore, the provision of mere technical details, which are not understandable for a lay person, will not suffice to satisfy this duty, the information should enable the subject to make use of the GDPR rights [9]. Hence it is a question of balancing expectations: while a detailed model description is not required, information must not be simplified to an extent that makes it worthless for the data subject.

Summary. To summarise, there are (broadly) three possible scenarios:

1. *Scenario 1:* If a medical AI system does not fall in the category "based solely on automated processing" (e.g. decision-support), Arts. 13–15 (duty to

provide information) and Art. 22 GDPR (limitation on automated processing) are not applicable (although providing information is recommended as good practice).
2. *Scenario 2:* If the medical AI system does per se fall under the prohibition of Art. 22 para 1 GDPR, its use is only permissible if it falls under one of the exceptions of this prohibition, the most important for medical AI being explicit (informed) consent. Even if explicit consent is given, the fully automated processing of special categories of data (including health data) requires additional safeguards, the bare minimum being the possibility of human intervention, the expression of the data subject's point of view and the possible contestation of the decision (Art. 22 paras 3 and 4 GDPR). Finally, it is obligatory to provide the necessary information according to Arts. 13 to 15 GDPR (1. The data subject (patient) is informed that automated decision-making takes place 2. provision of meaningful information about the logic involved and 3. explanation of the envisaged significance and consequences of the processing).
3. *Scenario 3:* If an AI system is fully automated and hence falls under the prohibition of Art. 22 para 1 GDPR, and there is no exception applicable, this type of AI system is illegal under the regime of the GDPR. This result confirms that the inclusion of a "human in the loop" is a key design criterion necessary for compliance with European "AI law".

2.3 Additional GDPR Requirements: Privacy by Design and Data Protection Impact Assessment

As stated above, transparency is a basic requirement of the GDPR. The guarantees we have just described are just one aspect of transparency. The GDPR also tries to foster an environment where better AI systems are developed already from the design-stage onwards, e.g. through risk based accountability (Art. 24 GDPR: technical and organisational countermeasures corresponding to the scope and risk of the intended processing) in combination with a privacy by design (Art. 25 GDPR) approach to ensure that data protection issues are already parts of the design and planning process and integrated directly into the data processing (e.g. pseudonymisation, technical transparency measures etc.) [8].

Before the implementation of medical AI, a so-called data protection impact assessment (DPIA; Art. 35 GDPR) will have to be provided: If processing (in particular by means of new risky technologies like AI) is likely to result in a high risk to the rights and freedoms of natural persons, a prior assessment has to be carried out. This assessment must contain a systematic description of the envisaged processing operations and the purposes (e.g. medical AI based on neural networks for melanoma diagnosis), an assessment of the necessity and proportionally of the processing in relation to the envisaged purposes (e.g. processing of health data for the purposes of treatment), and an assessment of the risk to the rights and freedoms of data subjects (risk of harming physical/mental health, privacy etc.). Furthermore, a description of the measures to address these risks (e.g. the use of XAI, restriction of features, anonymization etc.) is required.

Lastly, a possibility for the persons who are affected by the processing operations to provide feedback shall be established. As a matter of good practice, a DPIA should be continuously reviewed and regularly re-assessed. While there is no obligation to publish the DPIA, parts of it (e.g. a summary) should be accessible to the public (and especially to the data subjects) in order to foster trust and transparency [1].

2.4 Product Safety Law

The concept of privacy by design as well as the general assessment needed for a DPIA can also be thematically linked with general product safety procedures which are another key component of the EU secondary law framework for (medical) AI. In both cases, prior assessment is required before the implementation or the market approval of AI systems in order to secure their compliance with basic safety standards. While the DPIA is focused on risks for privacy, product safety regulations want to minimise the risk of harm by a faulty product, i.e. aim at securing a high level of safety for goods. Product safety and product liability provisions function as two complementary regimes, therefore these two mechanisms will be disscussed discussed in the following Sects. 2.4 and 2.5 [21].

Most medical AI applications will, if they are intended by the manufacturer to be used for human beings for specific medical purposes, qualify as medical devices under EU law. Product approval procedures for medical devices are in a state of transition and will soon be fully harmonized by two European regulations (Regulations 2017/745 and 2017/746). While the regulations were intended to enter into force by May 2020/2022, the Commission has recently proposed to postpone this date to May 2021/2022 due to the COVID-19-pandemic. Both regulations aim at securing a uniform standard for quality and safety for medical products to reduce barriers for entry for such devices caused by divergent national legislation. However, even these recent EU regulations do, somewhat surprisingly, not really address medical AI specifically and do not include AI as a special product category. However, most medical AI applications qualify as software intended by the manufacturer to be used for human beings for specific medical purposes and hence as a medical device (or part of a medical device). Based on the risk classification of these applications, conformity assessment procedures may be required (certification, review, clinical evidence to prove accuracy and reliability etc.). It has been noted that - at the current moment - the definition of software and its associated risk is relatively inflexible and does not differentiate between static and machine learning systems, thereby failing to address specific risks of ML (explainability, dynamic nature, false positives/negatives etc. [50]). In the USA the FDA has recognized this dynamic nature and the opacity of ML as sources of potential problems and presented a vision of appropriately tailored regulatory oversight to control the specific risks of ML [25,45] which the EU still lacks (though, as the above-mentioned White Paper shows [22], it is aware of this problem).

Medical AI used as therapeutic or diagnostic tool is at least associated with a medium (potential) risk and will therefore in any cases require market access

approval ("CE-marking") granted by private companies (so-called notified bodies), which is followed by a post market-entry assessment by national authorities (e.g. through collection and assessment of risk data by means of audits). To gain a CE-marking, (medical) devices must conform with the general safety requirements (mitigation of risks and a positive balance of benefit over risk). Explainability is not a core aspect in the assessment of general safety and performance (through a clinical/performance evaluation), but transparency of the system may be necessary to sufficiently demonstrate that it does not mistake mere correlations within data for causality [50].

Software should be designed according to the state of the art, which is specified by so-called harmonised standards. Verification and validation procedures are part of these standards, they require proper data management (bias avoidance), assurance of accuracy, ability to generalize etc. Again, explainability is not a specific part of this process, but it could be important for proper risk management, therefore as the PHG Foundation states convincingly: "Verification and validation may require some machine learning models to be made somewhat intelligible" [50]. Besides product safety law, it has also been convincingly argued that some degree of explainability of medical AI may also be required to avoid liability [30]; we will return to this later when assessing liability law. Because of the specific dangers of ML systems, continuous risk assessment (post-market surveillance, periodical safety reviews) by notified bodies have been proposed as a necessity if there is a substantial change to the product [37,50].

Future Developments in Product Safety Law. The need to properly address the risks of AI as medical devices has been recognized by the European Commission and the American FDA. The American FDA proposed that ML systems, which are not "locked" but continue to adapt, will require constant monitoring through their life-cycle, including additional review of modifications, and states that "Transparency about the function and modifications of medical devices is a key aspect of their safety." [25]

This assessment is shared by the Commission in the above-mentioned report [21] which enumerates the specific risks of AI which need to be addressed by the new legal framework. Among these risks are connectivity, complexity and the "autonomy" of AI-systems (i.e. the ability to learn) which could - when future outcomes cannot be determined in advance - make re-assessment during the life-cycle of the product necessary. Requirements for human oversight throughout the life-cycle of an AI-system and data quality standards have also been announced as part of planned new EU legislation. The Commission Report also addresses the problem of algorithmic opacity and states that product safety legislation does not explicitly address this risk at the moment, therefore making it necessary to implement transparency requirements (as well as requirements for robustness, accountability, human oversight and unbiased outcomes) to build trust in AI applications (e.g. by an obligation to disclose the design parameters and metadata of datasets). In the words of the Commission: "Humans may not need to understand every single step of the decision making process, but as AI

algorithms grow more advanced and are deployed into critical domains, it is decisive that humans can be able to understand how the algorithmic decisions of the system have been reached." [21]

2.5 Liability

While law is generally (relatively) flexible when addressing new technologies like AI, there seems to be one important blindspot: Questions of liability law. Although liability law is generally ambiguous by design because it has to cover many different scenarios, it remains silent in relation to AI, therefore it lacks necessary considerations, which leads to considerable legal uncertainty. Pertinent questions particularly unsettle the AI community. Who will be legally responsible when medical AI malfunctions? The software developer, the manufacturer, the maintenance people, the IT provider, the hospital, the clinician? Even a short analysis shows that many questions about "Who is liable?" and how can liability be proved cannot be answered conclusively or satisfactory by the current legal doctrine. While it would be an option to leave the questions to be answered by the courts and their case-law, this would nevertheless create considerable legal uncertainty (including differences between individual EU Member States) for at least some period of time. This is why the European Commission, for good reasons, plans to fill the void with a new and uniform European legal framework on (civil, not criminal) liability for AI.

European liability law broadly consists of national, non-harmonized civil liability and harmonized product liability law. Both these liability regimes have many ambiguities and problems when addressing AI. Civil liability (based on a contract, e.g. a medical treatment contract or on tort [=non-contractual liability]) is mostly fault-based, therefore normally the fault of the liable person, the damage and the causality between the fault and the damage must be proved [21]. Besides fault-based liability there is also strict-liability in specified areas (e.g. use of dangerous objects like motor vehicles): liability for a risk is attributed to a specific person by law (e.g. holder of a car) and the proof of fault (or causality between fault and damage) is not necessary. One idea behind strict liability is that while the use of the dangerous object is socially acceptable or even desirable, its operator should bear the liability irrespective of any fault on their part [39]. Fault-based and strict liability often overlap and function in parallel. With regards to the complexity and opacity of ML algorithms, it is questionable whether the burden of proof should lie with the victim. It will be hard to trace the damage back to human behaviour and to establish a causal link [21] (discussing this problem with regard to the national law of Austria [52], Germany [14] or the UK [53]). The concept of burden of proof therefore has been described as an nearly insurmountable hurdle. Due to this, many lawyers argue that when using AI the operator and/or the producer (e.g. the medical service provider [physician, hospital etc.]/designer) should carry the burden of proof [53]. Others go further than this and propose the introduction of a strict liability regime for AI [56,60]. The often mentioned possible downside to this strict liability approach is, that it could stifle innovation [5,14].

Liability for Treatment Errors. According to medical malpractice law a medical service provider (e.g. physician, hospital) could be held liable for treatment errors or for the absence of "informed consent". Normally, the health service provider is not responsible for the success of the treatment, only for providing a professional treatment according to the due standard of care which must be in accordance with current medical scientific knowledge (see above in the section on fundamental rights). Therefore, liability could for example be established if a doctor - who is by professional standards required to independently assess an AI recommendation using his expertise - realizes that the AI recommendation is incorrect but still bases a medical decision on it. However, the opacity of ML could it make it hard to assess 1. whether it was a reasonable decision to use AI, 2. whether the doctor was right or wrong (in the sense of adhering to the required medical standard) to deviate from an AI recommendation and 3. whether he hence is liable or not [47,53]. However, liability law not only addresses the (mis-)use of AI, it is also relevant as to the non-use of AI: If AI-assisted medical treatment reached higher accuracy than treatments without the involvement of AI, its use would constitute the (new) "due standard of medical care", making health care providers liable if they do not use AI for treatment or diagnosis [30,51,53]. Such an obligation cannot only result from liability law, but also from the European fundamental rights framework (see Sect. 1.4 above).

"Informed Consent". We already concluded that the lack of "informed consent" can also lead to liability. Consequently, the duties to provide information and to respect the autonomy of the patient are an integral part of contemporary medicine (and medical law). Already at the level of fundamental rights, Art. 3 para 2(a) CFR states that "informed consent" must be respected. This results in the concept of a "shared decision-making" by doctor and patient where the patient has the ultimate say, hence the patient and the doctor are seen as equal partners. The patient must have the autonomy to make a free informed decision, this implicitly requires sufficient information. This duty to provide information about the use and possible malfunctions of medical AI will depend on the risk associated with a particular AI system [47]. This implies that the patient must not necessarily be informed about each particular use of medical AI. It is interesting to note that a comparable approach can also be found in the American legal literature. While it has been argued, that there is no general duty to disclose the use of medical AI to patients, two factors have been identified which could establish an obligation to disclose its use to the patient: first, the opacity and second, the (increased) operational risk of an AI system [12].

The relation between "informed consent" and "explainability" is already being intensively discussed in legal literature. Various opinions on this question are expressed: Rather pragmatically, some authors point out that medical processes with and without the use of AI already reach a complexity today that obliges medical personnel to explain with "appropriately reduced complexity" [16,17]. Consequently, the idea of "informed consent" does not require that an AI decision is comprehensible in detail or that an explanation is given how a

specific decision has been reached through internal processes. According to this approach, it is sufficient if information is given in rough terms on how a medical AI application works and which type of mistakes can occur during its operation. It is interesting to note that there is a certain congruency between this (civil law) approach and the rather restrained reading of a "right to explanation" for automated processing of data under the GDPR (see Sect. 2.2. above). In other words: If health care providers use approved AI systems with a critical eye and point out possible errors of the system to patients, then they should – in case the patient has given consent - not be liable for errors that occur nevertheless (at best, the manufacturer might be liable [53]). Other opinions assume that medical personnel must be fully satisfied with the functioning of AI-based systems and otherwise must not use them (see note [51] where the author argues for risk-based validation; similarly, albeit more cautiously [47]) - which, consequently, should also enable them to inform patients comprehensively about the functioning of the AI-based system. According to this approach, medical AI can only be used if patients have been informed about its essential functions beforehand – admittedly in an intelligible form. These conflicting opinions clearly show that a legal regulation of this aspect would indeed be advisable to ensure legal certainty for health service providers.

Informed consent is a subject-matter which also shows the difference between "law in the books" and "law in practice". In real life, a patient's choice of a health care provider and a treatment method will often mainly depend on highly subjective aspects (e.g. the institution, the manufacturer of an AI system or the doctor is perceived as trustworthy) and not on the objective content of the information provided. However, as personal impressions can be misleading, the provision of scientifically sound facts as a base for informed decision-making remains, from a legal point of view, a cornerstone of the concept of "informed consent". Admittedly, this somewhat idealized concept of "informed consent" has sometimes been characterized as unnecessary complex and practically unattainable by health care providers. However, this criticism is not limited to the use of AI in medicine but part of the general discussion about the pros and cons of "informed consent" in medicine [16,17]. Furthermore, "informed consent" is not a monolithic concept, but has different manifestations. Routinely, the necessary information has to be provided by the health care provider in a personal consultation, however, in case of medication a package insert is (in addition to a prescription) regarded as satisfactory from a legal point of view. Building on this approach, the provision of a comparable "AI package insert" has also been sometimes been proposed as a means to meet the necessary transparency standards for AI [61].

Product Liability. Harm caused by a defective medical device is normally also addressed by European product liability legislation (in form of a directive), which holds the producers/designer liable for defects of the product even without fault. However, product liability primarily addresses "tangible" products (e.g. embedded software as part of a composite product), not "intangibles" like non-embedded standalone software [49,53]. Therefore, medical AI - as long

as it is not a composite part of a physical product - will be outside the scope of current product liability of manufacturers (this mirrors the situation in the American legal doctrine [32]). Furthermore, current European product liability law addresses neither AI's ability of continued learning and the resulting regular modifications of its models in a satisfying way as it focuses on the time of placing on the market and does not cover subsequent errors. It can be very challenging to prove that the defect of a ML algorithm had already been present at this point in time and could also have been detected [47].

Future Developments in Liability Law. Comparable to the legal framework on product safety procedures, the European Commission is well aware of the severe challenges in regard to civil and product liability for AI applications and announced several (including legislative) measures in the above-mentioned report on liability [21]. This should also contribute to a more harmonized legal regime across all EU Member States. In civil liability law the burden of proof concerning causation and fault will probably be adapted to mitigate the complexity of AI (e.g. shift of burden of proof from the patient to the doctor/health care provider concerning damage caused by medical AI).

With regards to product liability law the Commission will in the near future evaluate the introduction of a strict liability system, combined with compulsory insurance for particularly hazardous AI applications (that is, systems which may cause significant harm to life, health and property, and/or expose the public at large to risks). Such a system will presumably cover most medical AI applications - the use of which would also be affected by the proposed changes to the rules on the burden of proof. This new legal framework will probably also further address the difficult question of software as a product or as a service as the Commission has already announced that a clarification of the Product Liability Directive in this respect will be necessary [21,24]. Furthermore, the important concept of placing on the market as the reference point for liability could be changed to take account of the adaptability of ML; after-market assessment and monitoring therefore could become part of liability law [54]. Such an enhanced dual liability system (civil and product liability) would certainly help to eliminate many existing ambiguities regarding the liability of medical AI applications.

3 Conclusion

The European Commission stated in its White Paper [22] that "[...] while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors." As shown in this paper, this conclusion does not fully hold true for medical AI. As medical AI is closely intertwined with questions of fundamental rights, data protection and autonomy, it is a field of AI where the current state of legislation provides more answers to open questions than in other areas of the application of artificial intelligence. It must not be forgotten that legal provisions are

a technologically-neutral instrument, which can also (at least to a considerable extent) be applied to the use of (medical) AI even if it was not enacted with AI in mind.

As to the use of medical AI, fundamental rights and the GDPR both prescribe clear duties to provide information to enable "informed consent" and also require a "human in the loop" as an essential element of oversight. Even if it is doubtful that there is an outright "right to an explanation" of specific decisions, the patient must be provided with enough information to understand the pros and cons associated with medical AI and can therefore participate in the shared-decision-making process or make use of his privacy rights stated in the GDPR (e.g. Art. 22 para 3 GDPR). European AI must be developed and operated in accordance with the requirements of fundamental rights, including the protection of data and privacy (Arts. 8 and 7 CFR). We conclude: European medical AI requires human oversight and explainability.

We also showed that the current framework for product approval procedures and liability for AI has not so far addressed the use of AI in a satisfactory way, however, the EU took note of these inadequacies and plans to amend its legislation. In this regard, the Commission has announced that it will soon provide clarity by proposing market approval procedures which address the specific risks of AI (e.g. need for constant re-assessment) and a strict liability approach combined with an obligatory insurance scheme for malfunctions of AI. These proposed changes will also strengthen the focus on transparency. Furthermore, new legislation will possibly bring along changes in the burden of proof for product approval procedures and in case of damages allegedly caused by AI. Hence the new regulation will probably nudge AI developers to use explainable AI (XAI) to comply with its requirements more easily [30].

This European approach which is already visible in current legislation will be further substantiated by the work of the Commission which - in contrast to other nations - expressly pursues a "human-centric" approach to AI, which should be integrated in an "ecosystem of trust" [22], thereby necessitating transparency of AI applications. Therefore, the future of AI envisioned by the EU isn't a future where humans are mere cogs in a mechanical decision-making machinery, but a vision where AI still remains an important tool - but a tool only - to shape a future for humans. This seems particular relevant for medical AI which in many ways touches upon the very essence of a human being. In this respect, exaggerated fears of "Dr. Robot" [4] are not appropriate: The European legal framework for medical AI basically requires the use of explainable AI in medicine, thereby being well in line recent medical research on XAI [36,48].

Acknowledgements. The authors declare that there are no conflicts of interests and the work does not raise any ethical issues. Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554 "A reference model of explainable Artificial Intelligence for the Medical Domain".

References

1. Article 29 Data Protection Working Group: Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP248rev.01 (2017). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
2. Article 29 Data Protection Working Group: Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP251rev.01 (2018). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053
3. Article 29 Data Protection Working Group: Guidelines on transparency under Regulation 2016/679, WP260.rev.01 (2018). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
4. Bambauer, J.R.: Dr. Robot. UC Davis Law Rev. **51**, 383–398 (2017)
5. Bathaee, Y.: The artificial intelligence black box and the failure of intent and causation. Harv. J. Law Technol. **31**, 889–938 (2018)
6. Brkan, M.: Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. Int. J. Law Inf. Technol. **27**, 91–121 (2019). https://doi.org/10.1093/ijlit/eay017
7. Brkan, M., Bonnet, G.: Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: of black boxes, white boxes and Fata Morganas. Eur. J. Risk Regul. **11**, 18–50 (2020). https://doi.org/10.1017/err.2020.10
8. Bygrave, L.: Data protection by design and by default: deciphering the EU's legislative requirements. Oslo Law Rev. **4**, 105–120 (2017). https://doi.org/10.18261/issn.2387-3299-2017-02-03
9. Bygrave, L.: Minding the machine v2.0. The EU general data protection regulation and automated decision-making. In: Yeung, K., Lodge, M. (eds.) Algorithmic Regulation, pp. 248–262. Oxford University Press, Oxford (2019). https://doi.org/10.1093/oso/9780198838494.001.0001
10. Bygrave, L.: Article 22. In: Kuner, C., Bygrave, L., Docksey, C., Drechsler, L. (eds.) The EU General Data Protection Regulation (GDPR). A Commentary. Oxford University Press, Oxford (2020)
11. Casey, B., Farhangi, A., Vogl, R.: Rethinking explainable machines: the GDPR's 'right to explanation' debate and the rise of algorithmic audits in enterprise. Berkeley Technol. Law J. **34**, 143–188 (2019)
12. Cohen, I.G.: Informed consent and medical artificial intelligence: what to tell the patient? Georgetown Law J. (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529576
13. Datenethikkommission: Gutachten der Datenethikkommission (2019). https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html
14. Denga, M.: Deliktische Haftung für künstliche Intelligenz. Computer und Recht **34**, 69–78 (2018). https://doi.org/10.9785/cr-2018-0203
15. Dupré, C.: Article 1. In: Peers, S., Hervey, T., Kenner, J., Ward, A. (eds.) The EU Charter of Fundamental Rights. A Commentary. C.H. Beck - Hart - Nomos, Baden-Baden - München - Oxford (2014). https://doi.org/10.5771/9783845259055
16. Eberbach, W.: Wird die ärztliche Aufklärung zur Fiktion? (Teil 1). Medizinrecht **37**, 1–10 (2019). https://doi.org/10.1007/s00350-018-5120-8
17. Eberbach, W.: Wird die ärztliche Aufklärung zur Fiktion? (Teil 2). Medizinrecht **37**, 111–117 (2019). https://doi.org/10.1007/s00350-019-5147-5

18. Edwards, L., Veale, M.: Slave to the algorithm? Why a 'right to explanation' is probably not the remedy you are looking for. Duke Law Technol. Rev. **16**, 18–84 (2017)
19. Edwards, L., Veale, M.: Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? IEEE Secur. Priv. **16**, 46–54 (2018). https://doi.org/10.1109/MSP.2018.2701152
20. Etzioni, A., Etzioni, O.: Designing AI systems that obey our laws and values. Commun. ACM **59**, 29–31 (2016). https://doi.org/10.1145/2955091
21. European Commission: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (2020). https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf
22. European Commission: White Paper On Artificial Intelligence - A European approach to excellence and trust (2020). https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
23. European Data Protection Board: Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1 (2020). https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf
24. Expert Group on Liability and New Technologies - New Technologies Formation: Liability for artificial intelligence and other emerging digital technologies (2019). https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608
25. FDA: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback (2019). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
26. Fosch Villaronga, E., Kieseberg, P., Li, T.: Humans forget, machines remember: artificial intelligence and the right to be forgotten. Comput. Law Secur. Rev. **34**, 304–313 (2018). https://doi.org/10.1016/j.clsr.2017.08.007
27. FRA: Data quality and artificial intelligence - mitigating bias and error to protect fundamental rights (2019). https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf
28. Goodman, P., Flaxman, S.: European Union regulations on algorithmic decision-making and a "right to explanation". AI Mag. **38**, 50–57 (2017). https://doi.org/10.1609/aimag.v38i3.2741
29. Hacker, P.: Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. Common Market Law Rev. **55**, 1143–1186 (2018). https://kluwerlawonline.com/JournalArticle/Common+Market+Law+Review/55.4/COLA2018095
30. Hacker, P., Krestel, R., Grundmann, S., Naumann, F.: Explainable AI under contract and tort law: legal incentives and technical challenges. Artif. Intell. Law **16** (2020). https://doi.org/10.1007/s10506-020-09260-6
31. Haidinger, V.: Art 22 DSGVO. In: Knyrim, R. (ed.) Der DatKomm Praxiskommentar zum Datenschutzrecht - DSGVO und DSG. Manz, Wien, rdb.at (2018)
32. Harned, Z., Lungren, M.P., Rajpurkar, P.: Machine vision, medical AI, and malpractice. Harv. J. Law Technol. Digest (2019). https://jolt.law.harvard.edu/digest/machine-vision-medical-ai-and-malpractice
33. Harris, D., O'Boyle, M., Bates, E., Buckley, C.: Law of the European Convention on Human Rights, 4th edn. Oxford University Press, Oxford (2018)

34. High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for trustworthy AI (2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
35. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. **3**, 119–131 (2016). https://doi.org/10.1007/s40708-016-0042-6
36. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **59**, 29–31 (2019). https://doi.org/10.1002/widm.1312
37. Jabri, S.: Artificial intelligence and healthcare: products and procedures. In: Wischmeyer, T., Rademacher, T. (eds.) Regulating Artificial Intelligence, pp. 307–335. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-32361-5_14
38. Kaminski, M.E.: The right to explanation, explained. Berkeley Technol. Law J. **34**, 189–218 (2019). https://doi.org/10.15779/Z38TD9N83H
39. Koziol, H.: Comparative conclusions. In: Koziol, H. (ed.) Basic Questions of Tort Law from a Comparative Perspective, pp. 685–838. Jan Sramek Verlag, Vienna (2015)
40. Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., Müller, K.R.: Unmasking Clever Hans predictors and assessing what machines really learn. Nat. Commun. **10**(1) (2019). https://doi.org/10.1038/s41467-019-08987-4
41. Lipton, Z.C.: The mythos of model interpretability. ACM Queue **16**, 1–27 (2018). https://doi.org/10.1145/3236386.3241340
42. Malgieri, G., Comandé, G.: Why a right to legibility of automated decision-making exists in the general data protection regulation. Int. Data Priv. Law **7**, 243–265 (2017). https://doi.org/10.1093/idpl/ipx019
43. Mendoza, I., Bygrave, L.: The right not to be subject to automated decisions based on profiling. In: Synodinou, T.E., Jougleux, P., Markou, C., Prastitou, T. (eds.) EU Internet Law. Regulation and Enforcement, pp. 77–98. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64955-9_4
44. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019). https://doi.org/10.1016/j.artint.2018.07.007
45. Minssen, T., Gerke, S., Aboy, M., Price, N., Cohen, G.: Regulatory responses to medical machine learning. J. Law Biosci. 1–18 (2020). https://doi.org/10.1093/jlb/lsaa002
46. Mittelstadt, B., Russell, C., Wachter, S.: Explaining explanations in AI. In: FAT* 2019: Proceedings of the Conference on Fairness, Accountability, and Transparency, January 2019. pp. 279–288. ACM (2019). https://doi.org/10.1145/3287560.3287574
47. Molnár-Gábor, F.: Artificial intelligence in healthcare: doctors, patients and liabilities. In: Wischmeyer, T., Rademacher, T. (eds.) Regulating Artificial Intelligence, pp. 337–360. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-32361-5_15
48. O'Sullivan, S., et al.: Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. Int. J. Med. Robot. Comput. Assist. Surg. **15**, 1–12 (2019). https://doi.org/10.1002/rcs.1968
49. PHG Foundation: Legal liability for machine learning in healthcare (2018). https://www.phgfoundation.org/briefing/legal-liability-machine-learning-in-healthcare
50. PHG Foundation: Algorithms as medical devices (2019). https://www.phgfoundation.org/documents/algorithms-as-medical-devices.pdf

51. Price, N.W.: Medical malpractice and black box medicine. In: Cohen, G., Fernandez Lynch, H., Vayena, E., Gasser, U. (eds.) Big Data, Health Law and Bioethics, pp. 295–306. Cambridge University Press, Cambridge (2018). https://doi.org/10.1017/9781108147972
52. Reinisch, F.: Künstliche Intelligenz - Haftungsfragen 4.0. Österreichische Juristen-Zeitung, pp. 298–305 (2019)
53. Schönberger, D.: Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. Int. J. Law Inf. Technol. **27**, 171–203 (2019). https://doi.org/10.1093/ijlit/eaz004
54. Seehafer, A., Kohler, J.: Künstliche Intelligenz: Updates für das Produkthaftungsrecht? Europäische Zeitschrift für Wirtschaftsrecht **31**, 213–218 (2020)
55. Selbst, A.D., Powles, J.: Meaningful information and the right to explanation. Int. Data Priv. Law **7**, 233–242 (2017). https://doi.org/10.1093/idpl/ipx022
56. Spindler, G.: Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz - Braucht das Recht neue Haftungskategorien? Computer und Recht **31**, 766–776 (2015). https://doi.org/10.9785/cr-2015-1205
57. Topol, E.: Deep Medicine. Basic Books, New York (2019)
58. Wachter, S., Mittelstadt, B., Floridi, L.: Why a right to explanation of automated decision-making does not exist in the general data protection regulation. Int. Data Priv. Law **7**, 76–99 (2017). https://doi.org/10.1093/idpl/ipx005
59. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv. J. Law Technol. **31**, 841–887 (2018)
60. Zech, H.: Künstliche Intelligenz und Haftungsfragen. Zeitschrift für die gesamte Privatrechtswissenschaft **5**, 198–219 (2019)
61. Zweig, K.A.: Wo Maschinen irren können (2018). https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf