EUROPEAN LAW BLOG
NEWS AND COMMENTS ON EU LAW

ELB Blogpost 26/2023, 30 May 2023
TagTs: EHDS, AI Act, ChatGPT, LLM
Topics: Artificial Intelligence, Data Protection And Digital Governance, Public Health

The AI Act and European Health Data Space Proposal: Seeing 'AI to AI' With Each Other?

By Tjaša Petročnik, Sofia Palmieri and Jean-Aymeric Marot

Over the past few months, general-purpose artificial intelligence (AI) has emerged as a hot topic for policymakers. The 'AI hype' that followed the seemingly immediate success of the conversational bot ChatGPT has led to renewed calls for regulation in the EU, placing Large Language Models (LLMs) in the spotlight. Such models, which can be described as a subset of general-purpose AI that can process and generate human-like text, are increasingly being integrated into various industries. Healthcare is no exception, with potential applications ranging from clinical and operational decision-making to patient engagement. AI enthusiasts hope that LLMs will enable better communication between patients, physicians and healthcare workers, improve healthcare delivery or contribute to the advent of personal therapy bots. It remains to be seen which of these promises are feasible and realistic, but their capabilities have been and are continuing to grow exponentially, enabling them to reliably perform routine tasks for care providers like collecting information or navigating health record systems (an area of particular interest). Recent peer-reviewed studies have shown that virtual assistants and other AI-based tools can already perform many healthcare tasks better than specialist physicians, leaving experts wondering about the future of the human element in the patient experience.

However, generative AI applications in healthcare also raise important ethical and legal considerations that must be carefully addressed to ensure responsible deployment in such a sensitive context, in order to avoid harming patients and society, and ensure the new technologies are safe and trustworthy. Obvious concerns include the need for transparency and potential threats to data protection and privacy, as evidenced by the European Data Protection Board's recent announcement to launch a dedicated task force on ChatGPT. In addition, these models are known to 'hallucinate' quite frequently, producing inaccurate statements that sometimes go as far as contradicting the source content used to train them. In healthcare, inaccurate algorithm outputs can result in harm to health due to, for instance, a misdiagnosis or a recommendation of an inappropriate treatment. They also run the risk of inadvertently propagating stereotypes or other social prejudices against (already) vulnerable groups, which would lead to biased outputs as well as exacerbating disparities in access to and quality of healthcare. This particular point echoes the well-documented problems of under-representation of diverse populations in pharmacological and genomic studies, with very real implications for clinical practice.

With the use of AI-based technologies in healthcare likely to increase in the near future, not only due to technological development, but also as one of the proposed solutions to the rising costs of and demand for healthcare, it is necessary to examine to what extent can the (emerging) regulatory regime address the above-identified shortcomings and facilitate the development of trustworthy and safe AI systems. EU lawmakers are currently negotiating a final version of the much-awaited Regulation laying down harmonised rules on artificial intelligence (the 'AI Act'), while the European Commission has also put forward a proposal aiming to establish a set of rules and governance mechanisms to regulate the secondary use of health data in the context of a European Health Data Space ('EHDS').

Although the AI Act and the EHDS proposal are intended to fill legislative gaps, they are undoubtedly not taking place in a legal vacuum. In fact, they are very much in line with the EU's wider framework pertaining to artificial intelligence and health data. Notably, the Medical Device Regulation (MDR) aims to ensure the safety and performance of medical devices, including AI-based software, in healthcare settings. The upcoming AI Liability Directive tackles the issue of legal accountability when harm is caused by AI systems, whereas the General Data Protection Regulation provides a comprehensive legal framework for the collection, storage, and processing of personal data. While a deep dive into each of these instruments would fall beyond the scope of this blog post, in this contribution we highlight selected aspects of the interplay between these regulatory layers.

Risk management in the proposed AI Act

For those not familiar yet with the AI Act, this forthcoming EU regulation acts as a safety framework for AI systems. In brief, similarly to other safety regulations (such as the MDR) the AI Act proposes a classification based on the risks that a certain AI system might pose to health, safety, and fundamental rights. Depending on the risk class of the AI 'product', the manufacturer has to comply with different and progressively more demanding safety requirements.

The original version of the Act, dated April 2021, did not consider general-purpose AI. However, the version published in November 2022, after the feedback from the EU Council, took into account general-purpose AI, giving AI-insiders food for thought while ChatGPT made its way into the spotlight. According to Article 4(b) of the AI Act, general-purpose AI that might be used as high-risk AI shall comply with the requirements in place for high-risk AI usage (in Title III Chapter II of the Regulation). Based on Article 4(a), we understand that these requirements shall apply regardless of whether the system is placed on the market or put into service as a pre-trained model and whether further fine-tuning of the model is to be performed by the end user.

One might breathe a sigh of relief to know that safety requirements will also apply to general-purpose AI. Nonetheless, when applying the AI Act in the healthcare field, one might wonder how these shall be interpreted. As already explored in the literature, the applicability of the AI Act in the healthcare field follows the intertwining of the Act with the MDR. To simplify, the AI Act refers to the MDR, affirming that those systems that fall under the MDR and, according to the MDR, must undergo a third-party assessment, should be considered high-risk AI and therefore satisfy the requirements of the AI Act. For those not familiar with the MDR, this regulation is based on the identification of the device's 'intended use'. Based on the intended use of a device, the device could be a *medical device*. Furthermore, when the device is a medical device, the identification of the intended use is essential to classify the device into one of the MDR's risk classes. Due to the intertwining of the two regulations, the identification of the intended use of the AI is even more important. When discussing general-purpose AI, this whole architecture based on the identification of the intended purpose becomes rather fragile. In the absence of a pre-defined intended purpose, it is not possible to identify a specific AI system as a medical device, nor is it possible to identify in which risk class the AI-medical device falls both in the MDR and in the AI Act, thus upsetting the delicate balance between the MDR and the AI Act. Therefore, in the view of the authors, it seems an uphill battle to identify which general-purposes AI systems will be relevant to the healthcare sector and therefore must be considered high risk without reference to the MDR. This leaves manufacturers with the hard task of identifying *ex ante* which general—purpose AI might be used in healthcare. Shall they only refer indirectly to the definition of medical purposes in Article 2 of the MDR to identify high-risk medical uses of general--purpose AI? Or is the interpretative line even blurrier than the one drawn by the MDR? Some manufacturers might just take the 'easy road' in terms of interpretative effort, deciding to comply with the requirements regardless of the future intended use. This solution would not per se be a bad one, considering it would ensure general-purpose AI to be safe following the AI requirements, especially in the context of health. However, complying with the AI Act requirements, such as elaborating the technical documentation, might be a rather hard task for manufacturers of general-purpose AI. In this sense, the technical documentation following the AI Act requirements, requires foreseeing both the intended use of the system, which has to be included in the information provided to the users, and to foresee the risks that come with that specific intended purpose, which sounds like a contradiction in terms for general-purpose AI. All things considered, we might agree that it is quite difficult to comply with these requirements if the manufacturer still does not know the actual use that will be made of the general-purpose AI, leaving quite some uncertainty. Manufacturers could comply with these requirements by providing information about potential intended uses and general risks arising from general-purpose AI, irrespective of the specific application. This solution seems however only partial. On the one hand, general-purpose AI would be subjected to safety requirements ensuring, in a sense, the safety of the AI systems. On the other hand, the risk is that the compliance with the safety requirements of the AI Act will be only superficial, since it is not possible to identify for these particular AI systems an intended purpose in the first place. However, this problem might be mitigated by the inclusion of additional information once the intended purpose of the AI is finally identified. At that moment further information might be included in the technical documentation, as part of the continuous risk management mechanism put in place by the AI Act.

Further analysis of the matter could include the liability profiles dictated by the AI Liability Directive. Through a causality presumption, the Directive will hold the manufacturer liable for harms caused by the AI whenever it is possible to prove non-compliance with the requirements of the AI Act. Without delving into the intricacies of the Directive, we might just offer a brief reflection. The already rather complicated position of the manufacturer seems to get even worse when the AI Liability Directive is taken into account. On the one hand, we have the AI Act's requirement to provide technical documentation that is heavily dependent on the intended uses of the AI system. On the other, we have the near impossibility, by definition, of properly identifying the intended use for general-purpose AI. On top of this, the AI Liability Directive will hold manufacturers liable for any harm, even if it is not causally linked to them, if they fail to comply with AI Act requirements, in case of a harmful event. When it comes to general-purpose AI, this risk of compliance failure appears to be fairly realistic, which would expose manufacturers to significant liability claims.

Health data for AI: Shaping the European Health Data Space

To develop AI systems, such as the LLMs behind ChatGPT, having data is key. While healthcare is considered to be a data-rich sector, the reality is also that this data is generally under-used, including for the development and implementation of AI. In this light, facilitating the availability and (re-)use of health data for health-related research, policy-making, and innovation (the so-called *secondary use*) is one of the key objectives of the envisioned European Health Data Space (EHDS), in addition to empowering individuals with control over their health data in healthcare provision and fostering a digital health market. Building upon the GDPR, the EHDS will, say the Commission, act as 'a treasure trove for scientists, researchers, innovators and policy-makers working on the next life-saving treatment'. According to the EHDS Regulation proposal, published in May of last year, secondary use of health data – including data from electronic health records, registries, and medical devices as well as lifestyle data among others – should contribute to the 'general interest of society' and should only be possible for the purposes specified in Chapter IV of the proposed Regulation. One such allowed purpose in Article 34 is training, testing and evaluating of algorithms, while the Regulation also puts forward other purposes like scientific research, public health, health-related statistics, education activities, and providing personalised healthcare, among others. Conversely, secondary use is prohibited for activities that are detrimental or otherwise harmful to individuals or society at large. The proposal also introduces Health Data Access Bodies (HDABs) that will be in charge of granting access to health data for secondary use, by issuing a data permit. One of their many tasks specifically refers to supporting the development, training, testing, and validation of AI systems in the health domain under the proposed AI Act, reiterating the importance of the EHDS for affording the access to high quality health data for AI.

When considering the possible issues that might arise in relation to LLMs as AI systems that rely on health data to work, several come to mind. First, as is evident from the array of purposes comprised under the secondary use of health data put forward in the EHDS, these are conceptualised broadly and cover activities that are rather diverse in nature, like scientific research, providing personalised medicine, or development and innovation. As pointed out, the goals of the EHDS with regards to secondary use of health data will need to be appropriately balanced against data protection concerns. One of the criticisms that arise is that the above-identified secondary purposes are not properly delineated in the Proposal, meaning that they might fall under different grounds for exemption for processing of health data under the Article 9(2) GDPR. This itself might not be a problem if the HDABs that will issue permits for secondary use were provided with precise criteria on how to approach and assess such different types of use from the data protection perspective, but these are arguably lacking in the proposal that only requires HDABs evaluate whether the applications for secondary use of health data fulfil one of the allowed purposes, if the requested data is necessary for such purpose, and if requirements in Chapter IV of the Proposal are fulfilled. What is more, it has been rightfully pointed out that since the development of AI is its own separate purpose under the allowed secondary use in the EHDS, this might mean that the EHDS proposal could remove some 'health-related algorithmic or AI projects from the robust ethical and methodological standards that apply to *e.g.* scientific research'. This raises concerns about ensuring proper scrutiny (by the HDABs) over such endeavours when the data usage is not conducted by those engaging in scientific research, strictly speaking, but e.g. by private corporations that develop digital tools. While data-driven innovation is desired, we fear that a regime that is too permissive when it comes to data sharing might ultimately hamper health(care)'s fundamental values like safety or equity.

Next, one of the goals of the proposed EHDS Regulation is also to strengthen 'the rights of natural persons in relation to the availability and control of their electronic health data'. The approach of the EHDS is, in terms of the legal basis for secondary use, to not rely on a natural person's explicit consent. Further, since the HDABs do not have an obligation to provide information specific to each natural person concerning the use of their data, the question is whether the EHDS provisions provide for sufficient transparency for data subjects to exercise their rights and whether they even allow for natural persons to have an overview and a say in how 'their' health data is ultimately used. Crucially, the sectors from which the data for secondary use stems, including healthcare and health research, are particularly sensitive, also in terms of the trust placed in healthcare providers and researchers and the duties that define those relationships. It is thus of paramount importance that - in light of the possible issues associated with LLMs and general-purpose AI - facilitating secondary use of health data does not hamper this trust by e.g. leading to undesirable or even potentially harmful uses and AI applications, as this could have negative consequences for individual and population level health.

Conclusion

It is now rather widely accepted that LLMs (and other forms of AI) hold valuable promise for healthcare, but also that extra caution is needed given the sensitivity of this sector, which literally deals with matters of life and death. While the hype surrounding ChatGPT is new and (fairly) recent, the issues surrounding LLMs are certainly not. And this begs the question: is the EU's nascent approach to AI capable of adequately addressing these issues, or has the regulator been blind to the risks of harm posed by AI in the context of health? The answer is somewhat mixed: firstly, while it is reassuring that general-purpose AI that could be used as a high-risk system will have to meet the safety requirements set out in the AI Act, in practice it may be difficult for manufacturers to determine which general-purpose AI will be used in healthcare and for a medical purpose. This poses a compliance and liability challenge. Second, one could also raise concerns about access to data for AI development through the EHDS, and whether the regulator has provided the responsible bodies with a framework that ensures robust scrutiny of ethical, epistemological and methodological aspects of AI development for healthcare. In addition to ensuring a thriving internal market, we must not forget the protection of health as an essential component of social cohesion and social justice, and the potential of AI and AI more generally to threaten this. Regulators should not turn a blind eye to this.