RECIPES

Precaution • Innovation • Science

The use of AI in healthcare:

A focus on clinical decision support systems

The RECIPES project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824665

Authors

Tijs Sikma, Rathenau Institute

Rosanne Edelenbosch, Rathenau Institute

Petra Verhoef, Rathenau Institute

Contributors

Siebe Rozendal, IASS Potsdam

Sabrina Roettger-Wirtz & Ellen de Vos, Maastricht University

Harald Mieg, Humboldt University of Berlin

With thanks to:

Advisory board members

Linda Kool, Rathenau Institute

Manuscript completed in April, 2020

Document title	The use of AI in healthcare: A focus on Clinical decision support systems
Work Package	WP2
Document Type	Deliverable
Date	13 April 2020
Document Status	Draft for Advisory Board

Acknowledgments & Disclaimer

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824665.

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

Abstract

The aim of this case study is to better understand the complexities and controversies for applying the precautionary principle to the use of artificial intelligence (AI) in healthcare. The case thereby examines the theoretical applicability of the principle to a possible ‘emerging’ case, since there are relatively few examples of practical application of the precautionary principle to it. We focused on clinical decision support systems (CDSS). CDSS have historically been one of the main applications of AI in the medical domain and their risks are in many respects exemplary for the risks of the use of AI in healthcare in general.

Our analysis indicates that, in particular cases, the precautionary principle is theoretically applicable to the risks of CDSS. Though decision making in healthcare by humans is also accompanied by high risks, the implementation of CDSS pose *additional risks* because they often change the nature of the decision making itself. Depending on the scale of implementation and the type of decision, CDSS may harm individual health, public health and/or infringe on human rights. Moreover, there have been scientific analyses of these risks, but these analyses are characterized by a considerable amount of scientific uncertainty. This uncertainty is partially caused by the current lack of scholarship on this topic, but is also a consequence of ambiguities, complexities and uncertainties that are intrinsic to CDSS as a technology, the nature of healthcare environments and the types of risks concerned.

Our analysis of the EU risk governance shows that there have been precautionary warnings towards the necessary limits of decision making of AI in healthcare early on, that ‘precaution’ has been a standard for many CDSS developers and that a large collection of laws, regulations, norms and standards have emerged that partially cover the risks of CDSS. This, in addition to the fact that the precautionary principle originated from environmental law, may partially explain why the principle has not been applied to CDSS. Recently, the EU has moved from a more ethics/standards-based governance to a risk-based approach. In academic articles and public discussions there similarly is a shift visible towards a more precautionary approach towards the use of AI in healthcare, emphasizing the seriousness and uncertainties of especially data driven applications.

The precautionary principle may be useful for investigating the desirable limits of the implementation of CDSS. Policy makers, healthcare professionals and companies could ask themselves what the minimal requirements for a safe decision-making process in healthcare are and which decisions always should be ‘fully’ taken by humans. The principle might be instructive for reflexivity and awareness of the many uncertainties around the implementations of CDSS and could encourage anticipation, cocreation and incremental innovation, for which many possibilities exist in the innovation pathways of CDSS, as our study shows.

The innovation principle does not seem to be of relevance to this case. Careful consideration of the uncertainties and requirements of CDSS in the vulnerable domain of healthcare should have priority over the benefits of innovation in terms of jobs and economic growth or the health benefits that CDSS may offer on the long run. In many cases, moreover, it remains to be seen if the (partial) automation of decision making in healthcare is desirable and beneficial in the first place.

Table of Contents

List of abbreviations

CA	Consortium Agreement
CC	Consortium Committee
DOA	Description of Action
GA	Grant Agreement
PCG	Project Coordination Group
PO	Project Office
WP	Work Package

AI	Artificial Intelligence
CDSS	Clinical Decision Support Systems
GDPR	General Data Protection Regulation

1 Introduction

Introduction

In their daily practice healthcare professionals make decisions that have crucial consequences for the health and wellbeing of patients, or, in the case of communicable diseases, potential patients. Perhaps in no other domain in society is the decision over life and death so direct as in the domain of healthcare. Delegating this decision-making to machines – to clinical decision support systems – subsequently potentially brings forth direct risks towards (public) health and wellbeing.

The implementation of some CDSS moreover raises concerns with regard to human rights. The power over life that is exercised in healthcare is usually kept in check by a variety of procedures, standards and control mechanisms. A patient can for instance talk with a doctor about the decisions that are made over his body, the patient can check why these decisions are made, he/she can argue against them and he/she can trust that information about his body will not be used outside of the medical practice. Delegating decision-making in healthcare to machines can, as we will show in this case, put pressure on these assumptions. This may in extreme cases lead to violation of human rights, like access to healthcare, privacy, equality before the law and the autonomy over one's body.

A question that this case tries to answer is, do the concerns mentioned above sufficiently warrant the application of the precautionary principle? Are they serious, systemic and irreversible enough? And if so, when and how should the precautionary principle be applied? The other RECIPES cases draw lessons from how the precautionary principle has been applied in practice. This case instead examines the theoretical applicability of the principle to a possible 'emerging' case, since there are relatively few examples of practical application of the precautionary principle with regard to CDSS.

This case analysis therefore provides some unique lessons about the complexities and controversies surrounding the application of the precautionary principle on new technology. First of all, an analysis of why the precautionary principle is possibly applicable on the use of AI in healthcare forces us into reflection about what characterizes these risks in the first place. Secondly, a description of the risks, discussions about these risks and the their risk governance, illuminates controversies and complexities about why the precautionary principle is *not* applied. This case thus provides a glimpse of the theoretical and practical considerations made for not applying the principle.

To make the most of this analysis we have delimited our case ('The use of AI in healthcare') to a particular healthcare application: clinical decision support systems (CDSS). In many ways, as we will show, the use of AI in CDSS is exemplary for the general complexities and problems that surround the use of AI in healthcare. However, it should be noted that clinical decision support systems vary significantly with regard to their functionality and technical properties (see section 2). Therefore, we do not concentrate on one 'type' of CDSS technology but examine what precautionary considerations have been put forward with different types of CDSS, what risks appeared in relation to such systems and what this means for applying the precautionary principle or precaution in general.

Thirdly, in our analysis of the risk governance surrounding CDSS we focus on the European Union. The reasons for this are that, first of all, many of the RECIPES stakeholders operate on a European level, and secondly, that AI has recently become an urgent and major topic for policy makers in the EU. Lessons learned in this case may therefore be especially relevant and topical.

Key timeline

Political	*Legal*	*Science/risk assessment*	*Public debate*	*Other*

Year	Event	Relevance to case study
1942	Science fiction writer Isaac Asimov writes the Three Laws on Robotics	Asimov was one of the first to make an elaborate case for precaution towards thinking machines
1972	The development on MYCIN was started at Stanford University	MYCIN is often considered to be the first example of a clinical decision support system (a backward chaining expert system)
1976	Computer scientist Joseph Weizenbaum writes the book Computer Power and Human Reason	Weizenbaums book sparked off one of the first major debates in the AI research community about the preferable limits of AI.
2001 - present	A combination of technological developments – the rise of big data, the growth of sophisticated machine learning techniques and cloud computing – create large expectations with regard to the opportunities of AI	These developments were especially of importance for the capabilities of data driven CDSS
2017	The European Court of Justice decides that software can be seen as a medical instrument, even when the software does not have a direct effect on the human body	This made the use of AI in medical devices subject for CE marking
2017	The European Parliament adopts a resolution on the Civil Law Rules on Robotics	In the resolution the EP states that “Robotics research activities should be conducted in accordance with the precautionary principle, anticipating potential safety impacts of outcomes and taking due precautions, proportional to the level of protection, while encouraging progress for the benefit of society and the environment.’
2018	The European Commission presents its AI strategy	In the strategy the EU develops policies and risk governance more specifically focussed on AI
2018	The General Data Protection Regulation becomes enforceable	Many privacy risks of AI and CDSS are subsequently covered by the GDPR
2020	Publication of the White Paper on AI by the European Commission	The European Commission proposes a risk-based approach on AI and mentions the use of AI in healthcare as high risk. The EC asks for input from stakeholders

2 Clinical decision support systems

Clinical Decision Support System(s) (CDSS) are, in a broad sense, systems that support the decision making of healthcare professionals. Or to be more precise: 'active knowledge systems which use two or more items of patient data to generate case-specific advice.' (Wyatt and Spiegelhalter 1991). CDSS for example provide clinicians with alerts or reminders, highlight guidelines during care, provide suggested course of action and identify drug-drug interaction.

Their assistance is generally aimed at making the decision-making process for healthcare professionals easier, faster, less erroneous and more evidence based. The first CDSS were developed in the 1970's and the amount of different CDSS has grown enormously since then. Today, CDSS are often integrated with electronic health records and they sometimes make use of web-applications and/or are administered through a desktop, smartphone, tablet, biometric monitoring and wearable health technology (Sutton et al. 2020).

CDSS can differ significantly with regard to the type of medical practice they support. This can vary from administrative actions, for instance support to clinical coding and authorization procedures, to more medical procedures, such as plan processes, clinical diagnosis and condition-specific guidelines.[1] CDSS are used in both primary, secondary and tertiary healthcare. They are, for example, used by both general practitioners, specialists like cardiologists, and sometimes even by patients at home.

Because CDSS are so varied with regard to function and context of use, it is difficult to estimate how many people make use of them. Market research firm Reaction Data estimated in 2018 that 74% of healthcare organizations in the US make use of CDSS.[2] The market is dominated by large health IT firms like Cerner and EPIC.[3] According to BIS Research the global clinical decision support systems (CDSS) market generated a revenue of $1.57 billion in 2018 and is estimated to grow over $3.49 billion by the end of 2028.[4]

CDSS have historically been one of the main applications of AI technologies in the medical domain (Montani and Strianim 2019). Artificial Intelligence (AI), intelligence demonstrated by machines, is a core component of most CDSS. The support a CDSS can give to health care professionals is mostly based on the 'reasoning' that its AI provides. For instance, the CDSS's suggestion for a particular medical procedure follows from the comparison of data from the patient in question to the data in its system. According to its algorithms, the CDSS 'reasons' and comes to a particular advice.

In the context of the case study, it is important to note that there does not really exist 'one' type of CDSS technology. The properties and behaviour (and therefore the associated risks) of a CDSS are dependent on 'what kind of support' to 'what kind of decision making' they give.

First of all, the type of AI that a CDSS makes use of determines its capabilities and 'behaviour'. AI methodologies used for CDSS can be divided into two categories:

[1] OpenClinical, Decision Support Systems, http://www.openclinical.org/dss.html, last accessed, 15/6/2020.

[2] https://www.healthcareitnews.com/news/new-study-identifies-top-11-clinical-decision-support-vendors, last accessed, 15/6/2020.

[3] Cerner (25 percent), EPSi/Allscripts (14 percent), Epic (11 percent), Stanson Health (6 percent), Nuance (5 percent), Premier (5 percent), Truven/IBM (4 percent), Elsevier (4 percent), Zynx Health (3 percent), NDSC/Change (2 percent) and CPSI/Evident (2 percent). https://www.healthcareitnews.com/news/new-study-identifies-top-11-clinical-decision-support-vendors, last accessed, 15/6/2020.

[4] https://www.bloomberg.com/press-releases/2019-07-09/global-clinical-decision-support-systems-market-to-reach-3-49-billion-by-2028, last accessed, 15/6/2020.

knowledge-based AI and data driven AI (Montani and Strianim 2019). In the case of knowledge-based AI, a 'top down' attempt is made to model human knowledge in computational terms. These CDSS consist of a knowledge base, an inference engine (an 'if-then-structure'), and a mechanism to communicate. Medical diagnoses and the accompanying symptoms are for instance translated to the knowledge base and once someone consults the computer by typing in particular symptoms, the computer will show the corresponding diagnosis.

Data driven CDSS start 'bottom-up' and infer suggestions on the basis of the data that is fed to it, for instance a large amount of data about patients and the (correct) diagnoses that doctor has made. By linking variables, it learns to 'recognize' the patterns of appropriate diagnoses; which symptoms fit with which diagnoses. While in knowledge based CDSS the rules followed are coded by humans, a data driven CDSS 'finds' rules through the data. A data driven AI can therefore, in many cases,[5] not explain 'why' it follows a particular rule. Data driven CDSS can moreover be subdivided according to different types of machine learning techniques, like support-vector machines, artificial neural networks and genetic algorithms (Montani and Strianim 2019). The complexity of these types of machine learning can make them especially prone to high risks. They are more unpredictable than knowledge-based CDSS, and when something goes wrong, it is more difficult to find out 'what' goes wrong and how it can be fixed (see chapter 3 for examples).

Besides the type of AI, CDSS are categorized on the basis of system function (some systems advise on what is true/diagnose while others advise on what to do/the treatment), the model used for giving advice (passive or active), style of communication (consulting or critiquing), human computer interaction (for instance voice recognition or keyboard) and if they are used for pre-diagnosis, during diagnosis or post-diagnosis (Wasylewicz and Scheepers-Hoeks 2018).

Potential benefits of CDSS

Proponents of CDSS argue that CDSS improve the decision making in healthcare. The main argument is that the reasoning of the AI in a CDSS adds value to the overall decision-making process of healthcare systems (Verughese et al 2017). This supposed value is dependent on the specific place the CDSS gets in overall decision taking in healthcare.

Some CDSS are primarily developed to 'replace' or 'mimic' the existing reasoning of healthcare professionals. In these cases, the added value lies in the fact that the AI does the same as its human predecessor, **but faster, more accurate, with less costs and less 'human' errors**. An example would be a virtual nurse that automatically diagnoses the patient and prescribes medication through chat (or voice recognition).[6]

Secondly, there are also CDSS that are primarily developed to 'augment', human decision making. In this case, a healthcare practitioner makes use of particular capacities of the CDSS to **improve his decision making**; the added value here lies in how the machine complements human reasoning. Quick Medical Reference, for instance, augments the

[5] Due the emergence of the right to explanation, methods and techniques in the application of artificial intelligence technology are developed so that the results of the solution can be understood by human experts. This so-called 'Explainable artificial intelligence (XAI)' is still largely in the development phase and it is very much uncertain if explainability is feasible for all data driven AI applications.

[6] See for instance the 'virtual nurse'. https://www.careangel.com/ai-and-voice-powered-virtual-nurse-assistant. Automated medical/health advice is in a sense already prevalent in health apps: Niezen, M.G.H., Edelenbosch, R., Van Bodegom, L. & Verhoef, P. (2019). Health at the centre – Responsible data sharing in the digital society. The Hague: Rathenau Instituut.

ability of a doctor to diagnose patients with a knowledge base of diseases, diagnoses, findings, disease associations and lab information.[7]

Finally, in some cases a CDSS does not 'replace' or 'augment', but makes entirely new decisions possible. Data driven CDSS can for instance provide new information based on correlations between data sets that were unobservable before. In the so-called 'Learning Healthcare Systems' data driven CDSS play a role in finding new ways to continuously learn from data about the performance of the healthcare system and make improvements accordingly (Dagliati et al. 2018).

Taken together, these advantages suggest that CDSS can make healthcare more efficient and possibly more effective. The efficiency lies in a reduction of the costs, efforts and time that has to be invested in decision making. Time which health care professionals can invest in human contact with the patient. However, it should be noted that the presumption that CDSS will provide increased efficiency is often contested.[8] In some cases researchers argue that more long-term studies are needed to measure the added benefits (like decrease in deaths or medication errors) (Jia et al. 2016). Some suggest that the use of CDSS may also take up extra time, effort and costs to instruct personnel and to maintain the necessary infrastructure.[9] Also effectiveness has to be proven still in many cases (Moja et al. 2014; Murphy 2014).

A final potential benefit is the broad scope of AI applications that CDSS can indirectly contribute to. Because AI is a general-purpose technology, investment in research and development of CDSS might trickle down into progress in other domains where AI or related technologies are used. This might increase the technological competiveness of a country, the export of innovation to other countries and attract foreign capital (like investments or researchers) (Castro and McLaughlin 2019).

3 Risks and scientific uncertainties

Risk/threat

3.1.1 Potential risks

In their daily practice, healthcare professionals make decisions that have crucial consequences for the health and wellbeing of patients, or, in the case of communicable diseases, potential patients. The augmentation, replacement and supplementation of this

[7] Open Clinical, Decision Support Systems, http://www.openclinical.org/dss.html, last accessed, 15/6/2020.

[8] It is difficult to make general conclusions on effectiveness, since this largely depends on the CDSS used and, for example, the disease in question. One study identified six medical conditions, in which CDSS improved patient outcomes in a hospital setting. Another study stated that: 'There is a large gap between the postulated and empirically demonstrated benefits of [CDSS and other] eHealth technologies ... their cost-effectiveness has yet to be demonstrated'. See respectively: J. Varghese et al. (2018). "Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review". Journal of the American Medical Informatics Association. 25 (5): 593–602. A.D. Black. Et al. (2011). "The impact of ehealth on the quality and safety of health care: A systematic overview". PLOS Medicine. 8 (1).

[9] It is as of yet difficult, for instance, to continually adequately incorporate the extensive quantity of clinical research in such systems.

decision-making with CDSS is in this sense accompanied by risks for individual health, public health and human rights.[10]

It should be noted though that the decision making in healthcare is *always* accompanied by risks. There is always the risk that a doctor, intentionally or by accident, prescribes the wrong treatment. Many important medical decisions moreover necessarily have to be taken in the context of considerable (scientific) uncertainty.

What concerns us in this case, however, is the *additional* risks that CDSS pose. The introduction of a CDSS transforms how decisions are made in healthcare and therefore pose *new* risks. On the basis of a literature study[11] we found four ways in which CDSS transform the healthcare system and therefore pose additional risks: 1. Because they rely on data accumulation or datafication. 2. Because they imply a loss of human control. 3. Because a human element is removed in the decisions. 4. Because they imply a new division of labour and responsibilities in the healthcare domain.

Many of the risks of CDSS are concerned with the question of what ‘good’ decision making in healthcare entails and to what extent things like privacy and autonomy of the patient, transparency, accountability and reflexivity are necessary to ensure that the health of patients is served sufficiently. We will first give a broad overview of these risks. The extent of which the precautionary principle is deemed applicable, will be discussed in the section ’Relevance of the precautionary principle to the case’.

1. Risks related to datafication

First of all, the augmentation, supplementation and replacement of decision making by CDSS is dependent on data accumulation. A CDSS can only come to correct suggestions when important elements of its environment have been ‘translated’ into discrete data. A CDSS for instance reads the biometric data of a particular patient, compares this to the data it already has about symptoms and diseases and subsequently formulates a diagnosis. In the case of data driven AI, the algorithms also are formed on the basis of the data available to the CDSS. The use of CDSS is thus dependent on a datafication (and digitization) of the healthcare system. New risks emerge that are related to the dependency of CDSS on digital data.

First of all, medical information about an individual is, by its very nature, personal, intimate and sensitive. An individual’s right to privacy gives him/her a choice in whether he/she wants to disclose this information about himself/herself.[12] The risk of violating privacy is exacerbated when one takes into account genetic data or other data that not only informs about an individual, but also his family or environment. Moreover, correlations on the basis of biometric data may not always be self-evident; an iris can for instance show that someone has diabetes or high blood pressure, and irregularities in fingerprints may indicate leukaemia or breast cancer (Kool et al. 2017).

Secondly, when health-data and CDSS are used to produce and apply medical knowledge, this can change who decides on what constitutes health and disease. A CDSS that produces (medical) knowledge implies a delegation of this responsibility to the developers of these systems. The specific algorithms and data-sets these developers use to train the AI of an CDSS determines the knowledge that comes out. Especially when clinical support is used outside of the supervision of healthcare professionals, for instance in connection to health-

[10] The precautionary principle has been acknowledged by the European Court of Human Rights (EHRM) in relation to human rights. Tătar EHRM 27 januari 2009, ECLI:CE:ECHR:2009:0127JUD006702101 (Tătar/Roemenië). It should be noted though that the application of the principle in relation to human rights does not seems to be custom.

[11] For a full overview of the literature used, see ‘References’ in the back of this report.

[12] To be precise, this is about ‘Informational privacy’; the capacity of an individual to control information about himself/herself.

apps, this gives rise to new risks. Incapable developers may unknowingly prescribe wrong health-information and developers with ulterior motives could prescribe health information that benefits them or their client, like an insurance company. As such, this brings risks for doing harm in healthcare (Niezen et al. 2019).

There is also the risk of bias that may lead to suboptimal healthcare, in particular for vulnerable groups or women. Especially in the case of gender and sex, there exist substantial biases in existing medical data. Historically, norms and classifications in the medical sciences have predominantly been based on male bodies. It is presupposed that 'anatomy' is first of all the anatomy of the male. However, researchers have found sex differences in every tissue and organ system in the human body, as well as in 'the prevalence, course and severity' of the majority of common human diseases' (Perez 2019). They have even found differences in cells (Perez 2019). Existing biases may thus be prolonged and even exacerbated in CDSS. Feminist and journalist Caroline Criado Perez notes that: 'The introduction of AI to diagnostics seems to be accompanied by little to no acknowledgement of the well-documented and chronic gaps in medical data when it comes to women.' (Perez 2019). Machine learning can amplify such existing biases.

Thirdly, the datafication of health poses new risks when medical knowledge of someone is used to have power over someone. The ability to draw conclusions from diverse data sets might make people vulnerable to the extent that knowledge of their physical, emotional, social or psychological constitution is of interest to third parties like employers, health insurance companies, scammers, and competing football teams. Used in this way, CDSS could pose structural problems in relation to profiling and discrimination. Healthcare data has already been targeted by criminal organizations to be used by extortion or for long-term identity theft (Steger, 2019).

Fourthly, the datafication of healthcare can lead to new ways to manipulate people's behaviour. CDSS based information about biological constitutions, psychological predispositions and behavioural patterns can potentially be used to extrapolate, predict and therefore influence behaviour. This might result in asymmetries of power and information (Council of Europe, 2018) and conflict with, amongst others, the right to not be measured, analysed or coached.[13] All in all, this poses risks for the autonomy of a healthcare professional over his profession and the autonomy of a patient over his/her body and health.

2. Risks related to a loss of control

A substantial difference between decision-making by a human and decision-making with the help of a CDSS, is that with the latter a machine is (partially) in 'control'.

Aspects of control taken over from healthcare professionals by a CDSS may include decision making about what to examine, reasoning about observations or control over what is done with the results. In cases where the use of a CDSS has become habitual, it can replace the considerations a doctor would have about what to examine or to do. Moreover, the autonomy to decide about what to share with, for example, other departments may be limited when a CDSS is connected with others systems and automatically shares this information with other databases.

To the extent that the reasoning of the AI is a black box (Price 2015), it might wrongly give 'objective' standardized conclusions, even in situations that require a non-standard approach. Existing biases in medical knowledge that are translated into the algorithms

[13] Proposed by, amongst others, the Rathenau Institute. Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool (2017) Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality–Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE), The Hague, Rathenau Institute.

might unconsciously become normalized because a healthcare professional might just 'trust' whatever 'objective' output the computer provides. Institutionalized racism, genderism and sexism might however be reproduced in machine learning models.

A lack of control can also result in a lack of responsibility and accountability (Price 2015). It may become unclear who, why and how a decision was made. The blame of a mistake could for instance be attributed to the developers, implementers, healthcare professional, data supplier and/or system manager of a CDSS. Responsibility, accountability, explainability and transparency are however essential in the case of justifying and communicating on medical decisions, solving problems and preventing future mistakes.

Moreover, an overreliance on AI in medical problem-solving and decision making could result in the loss of appropriate skills and knowledge among health professionals (deskilling) (Gheeshan et al. 2009). In the case of a malfunction of the AI system, this gives rise to new vulnerabilities.

3. Risks related to the lack of a human element

Another substantial difference between a decision made by a CDSS and a human is that every cognitive act of a human, a 'human element' is directly present. When a healthcare professional 'thinks' about what to do, the whole of his 'humanity' is present: self-awareness, empathy, social intelligence, emotion and sincerity.

Delegating cognitive tasks to a machine essentially could mean removing these aspects from the decision-making process. Healthcare professionals make use of implicit knowledge and subtle skills that are sometimes difficult to formalize and make computable (Coeckelbergh, 2013). This could also remove aspects of 'care' from healthcare. Far-reaching automation might consequently endanger the 'right' to human contact or even the right to healthcare to the extent that care necessitates a person that 'cares for' or is 'involved with' your suffering when decisions are made.

Healthcare professionals moreover often have to make difficult decisions on the basis of conflicting research. Such careful deliberations and reflection (meta-analysis) are difficult or even impossible to translate into the reasoning of a CDSS (Gardner, 2004).

4. Risks related to another division of labour

The replacement of decision making in healthcare with CDSS tends to be accompanied by a new division of labour. Other actors, like IT companies and data collection agencies, acquire a (more important) place in the domain of healthcare (Niezen et al. 2019; Kobie 2019). This can bring forth new dependencies and therefore new risks.

When more processes are delegated to AI systems, the health care system becomes more dependent on those that develop, maintain and update these systems, handle the data and develop algorithms. The accumulated benefits of data can lead to monopolization in the data market. As a consequence of this, expertise and possession of data resources would rest in the hands of fewer companies, which could result in higher costs. This can put a severe strain on publicly funded healthcare.

Moreover, the processing of these data often happens outside the territory of the healthcare system itself, for instance in the cloud. This could mean that knowledge-production and factual expertise in this domain is increasingly in the hands of outside actors (Niezen et al. 2019). This can make health services more dependent and vulnerable, since they are not completely under control of the healthcare organization (the so-called lock in effect).

Scientific analysis

Some form of scientific analysis has already taken place with regard to the risks of clinical decision support systems. These analyses can be subdivided into analyses about a particular system (like IBM Watson), about a particular type of system (like data driven clinical support), about a particular type of risk (like data risks) or CDSS in general.

A quick literature scoping reveals that analyses have been made in in the field of AI research, computer science, (Bio)-ethics, STS/TA-institutes, Medicine, Health IT, Risk governance, risk assessment, Law and policy studies.[14] These analyses are often based on the experience and intuition of experts (what they expect could/would happen), informed reasoning and by collecting the perspectives of stakeholders.

To some degree, clinical trials have been executed on CDSS. In most instances these studies seem to focus on effectivity and economic benefits (Verughese et al. 2017), and there still exists considerable uncertainty about the long-term effects (Jia 2016) and the more ambiguous and complex risks (with regard to a loss of control, another division of labour, lack of a human element and data risks). This is possibly related to the fact that these trials are more focussed on technical and measurable effects, while these ambiguous risks may more often play a role on a management/policy level. For instance, a clinical trial might measure if a CDSS works appropriately, but it cannot (easily) say anything about the question if the use of the CDSS significantly reduces the autonomy of the healthcare professional or leads to risky dependencies on IT developers.

Many of the main risks of CDSS seem difficult to reduce to standard risk assessment procedures. Risks related to deskilling, deresponsibilization, data-abuse or the absence of humanity in medical decision-making are difficult to formalize and standardize, especially because such risks highly differ with regard to the type of CDSS and the environment in which it is used. A definitive body of work with robust (quantifiable, testable, repeatable etc) scientific statements about 'the risks of CDSS' seems to be absent, though a variety of tests and monitoring has been done about the effectiveness of CDSS in practice.

Scientific uncertainty

Some analysis has taken place in the scientific community with regard to the risks of CDSS (see previous paragraph). However, much of the work on the risks of CDSS is characterized by scientific uncertainty. Some degree of uncertainty seems to correlate with the status of current scholarship, which is fragmented and, with regard to new types of (data driven) CDSS, relatively new. Reasoning in most work on the risks of CDSS is mostly speculative and not based on large sets of empirical data.

The lack of scientific certainty and consensus surrounding the risks of CDSS is however also a consequence of some uncertainties inherent to the use of CDSS. First of all, because CDSS make use of AI, especially in the case of unsupervised machine learning, its behaviour and effects can be complex and difficult to predict. We call this '**technological variability**'.[15] Secondly, a CDSS always interacts with the complex and uncertain environment of a healthcare system. It is thus difficult to estimate if a particular CDSS will function adequately in line with the expectations, requirements and standards of the healthcare professionals. We call this **'environmental variability'**. Thirdly, the main risks that are concerned with CDSS (see section 3.1) are difficult to measure objectively. It is difficult, for instance, to measure and estimate the outcome and the chance of 'deskilling', when a CDSS is substantially biased or when a human perspective is needed for a decision. We call this **'risk assessment variability'**.

[14] See 'References' for an overview of the consulted literature.
[15] We define variability as a lack of consistency or fixed pattern.

We will analyse the properties of the risk variabilities with regard to **complexity**, **uncertainty** and **ambiguity** in the next sections. In each subsection we distinguish between variabilities caused by the nature of CDSS technology, the nature of the environment in which they are used and the type of risks concerned.

3.1.2 Complexity

Scientific uncertainty surrounding the risks of CDSS is partially a consequence of complexity in multiple ways.[16] Both the behaviour of CDSS, the environment in which they are used and the types of risks display properties of a complex system.

Complexity of the technology
CDSS that make use of machine learning, especially in the case of unsupervised machine learning, may display emergent and self-organizing behaviour.[17] Most CDSS do not yet make use of machine learning or are still in development, but future applications that do may exhibit the same types of complexity. A CDSS that makes use of machine learning which has the generic aim to support decision making could for instance try to optimize its support and combine data or develop algorithms of which a healthcare professional had not thought of.

Complexity of the environment
Not only the CDSS, but also the environments in which they are used, are characterized by complexity. When a CDSS is implemented in a healthcare system, for instance a hospital, it has to be attuned to a system that consists of many interacting elements. For a good application it has to be attuned to the expectations, existing norms and standards of healthcare professionals. The messages of a CDSS for instance have to be readable, understandable and helpful in the context of the daily tasks of a doctor, the specific needs of a patient and the oversight of a manager and/or a privacy officer.

The complexity of the behaviour of a CDSS can moreover become more extensive because it interacts and adapts to complex and unpredictable entities: humans. An AI system can therefore encounter many forms of reflexivity.[18] In the case of a CDSS this can mean that its algorithms change on the basis of the people that operate it and the people that constitute its database.

Another cause of complexity is that a CDSS sometimes has to mediate between different standards, inputs and multiple different sets of data. Interoperability of data is necessary for the development of good AI systems. Currently, the medical landscape is however characterized by a large number of disconnected small data from different technical systems (For instance: different electronic medical records, wearables, mobile health apps) that use different standards and protocols (Lehne et al. 2019; Niezen et al. 2019).

The interaction of a CDSS with other (AI) systems can in some cases moreover lead to feedback loops. In the US, for instance, a biased medical algorithm delayed healthcare for black people (Obermeyer 2019). The algorithm was used to predict the future health of individuals on the basis of their past health records. It identified people who were likely to

[16] See ‘WP2 Conceptual framework for comparative multiple case study analysis’ for an overview of our conceptualization of complexity.

[17] In the OpenAI project the AI players in a game were for example said to demonstrate ‘emergent behaviour’. They developed strategies that the developers had not thought of themselves. Strickland E. (2019) AI Agents Startle Researchers with Unexpected Hide-and-Seek Strategies, Institute of Electrical and Electronics Engineers.

[18] Reflexivity describes how human agents perceive, anticipate and alter the systems in which they are participating within the specific social, cultural and technological constraints being faced. This implies that by perceiving and acting in the system, individuals alter that very system in a type of dynamic feedback loop between the course of events and agent perceptions of those events. See RECIPES WP2 Conceptual framework for comparative multiple case study analysis.

need extra care in the future. For a variety of socioeconomic reasons related to access to healthcare, black patients make less use of healthcare and thereby generate lower costs than white patients. Subsequently the algorithm prioritizes white people over black people with the same health status. As a consequence of such biases, black people may tend to trust the decision making in healthcare less, which will again be reflected in the data (which will show that they apparently have less healthcare costs). Such biases are difficult to discover beforehand, because they often depend on unknown unknowns (an AI may for instance indirectly take into account the gender because of the wordings that are used), the developers of AI systems tend to ignore complex social contexts and because 'bias' and 'fairness' are in itself difficult and ambiguous notions (Hao 2019).

Complexity of risk assessment

The risks that we distinguished in section 3.1 are also characterized by complexity in multiple ways. To the extent that 'good' (and consequently safe) decision making in healthcare consists of many elements, so risks can be a consequence of multiple elements, which are itself complex. Good decision making may include respect for the privacy and autonomy of the patient, transparency, accountability and reflexivity. To the extent that a CDSS replaces, augments or supplements the decision making, it may impair the decision process with regard to each of these elements. However, what does sufficient privacy, autonomy, accountability or transparency for instance exactly entail and how should each of these elements be balanced with efficiency and effectiveness? These are complex questions.

Moreover, many of the main risks described in section 3.1 may be intertwined and their relation is difficult to assess. For instance, the risk that a healthcare system becomes overly dependent on the infrastructure and knowledge of outside actors, may pose risks related to data, loss of control or lack of human elements in the decision-making process. A commercially oriented actor may scrap human intervention as much as possible to spare costs, sell data to insurance companies and take away control from healthcare personnel to improve efficiency. But such interdependencies are very difficult to assess and predict.

3.1.3 Uncertainty

Uncertainty describes the lack of knowledge about the outcomes or likelihoods, or both, of an event or process.[19] Both the behaviour of CDSS, the context in which they are used and the risk assessment are characterized by uncertainty.

Uncertainty of the technology

In the case of machine learning, the learning capabilities of a CDSS can gives it some 'autonomy', which can make the impact uncertain: 'tasks performed by machine learning are difficult to predict beforehand (how a new input will be handled) or explained afterwards (how a particular decision was made).' (Mittelstadt et al. 2016). Moreover, to the extent that an AI system comes to conclusions on the basis of statistical inferences, its decision-making is always based on probabilities, and thus (partially) uncertain knowledge (Mittelstadt et al. 2016). Though, it should be noted, the same (often even to a larger extent) of course applies to human decision-making.

The behaviour of a CDSS can also exhibit uncertainty[20] in the sense that small variations in the initial conditions of a (learning) AI system (for instance: its core code statements) can have highly divergent results. Researchers warn for instance for cyberattacks that can change the behaviour of machine learning AI systems by using only tiny pieces of digital data. Changing a few pixels on a lung scan could for instance fool such a system into detecting a non-existing disease (Finlayson 2019).

[19] See 'WP2 Conceptual framework for comparative multiple case study analysis' for an overview of our conceptualization of complexity.

[20] 'Variability uncertainty arises because of relevant, correct, but 'random' system behaviour.' RECIPES WP2.1.

Epistemic uncertainty can follow from the fact that the design of an AI system or the way it is connected to other IT-systems can be obscure. This makes it harder to predict its consequences, and therefore: it's risk. IT systems depend on interoperability between different codes, protocols and applications. Especially in relation to older IT-infrastructure that are written in older programming languages, it can be difficult to ascertain how it will interact with new systems (see also 3.2.2.1 Complexity).

Uncertainty of the environment
The environment in which a CDSS is used – a healthcare system – may besides complexity also be characterized by uncertainty. Healthcare professionals often have to make decisions under uncertainty about events as well as the likelihood of these events (for instance in the case of an unknown disease). Subsequently, it can be difficult with regard to the implementation of an CDSS to take into account these uncertainties and predict the risks of a CDSS. A CDSS in principle has to be prepared for many situations, but to the extent that these uncertain situations occur a CDSS may be unsuitable (unbeknownst to personnel) and thereby potentially pose additional risks.

Due to the complexity of many healthcare systems (Panch et al. 2019) it is moreover difficult to test the likelihoods of uncertain outcomes in controlled trials; the dynamic of a healthcare system and the extent an AI systems fits is difficult to simulate realistically.

Uncertainty of risk assessment
Just like that the types of risks described in section 3.1. are complex, they are also uncertain. For instance, it is very difficult to predict what the consequences would be if large amounts of health data fall into the wrong hands both with regard to the outcomes as the likelihood. It is difficult to estimate to what extent such data sets can be traced back to individuals and to what extent or how it can be used against them. The combination of separate data sets can lead to unexpected conclusions (Kool et al. 2017). Of apparently innocent biometric information sensitive correlations may for instance be discovered with regard to biological aspects like heritable diseases, psychopathology, behavioural dispositions, preferences or pregnancy.

Similarly, it is difficult to assess the amount of harm that is caused by decisions made on (partially) defective data. It is for instance difficult to measure how many harm has occurred due to the fact that there exists a strong bias towards a particular male body in medical data (Perez 2019).

3.1.4 Ambiguity

Another cause for scientific uncertainty on the risks of CDSS is that they are characterized by interpretive[21] and normative[22] ambiguity. Both the behaviour of CDSS, the context in which they are used and the risk assessment are characterized by ambiguity. This also brings forth ambiguity with regard to what extent risks are present when a CDSS is implemented.

Ambiguity around the technology
First of all, ambiguity lingers about what AI exactly is and when a CDSS exactly makes use of it. AI is still a relatively open-ended notion about which diverse conceptualizations are used. No transnational agreement exists with a commonly accepted working definition, neither at the technical nor the legal/policy level (EPRS/STOA 2019). However in the EU there is some consensus on the policy level. There also seems to exist some ambiguity

[21] 'Interpretative ambiguity refers to the situation where information, data, analyses and risk governance strategies are interpreted in different ways by different actors.' RECIPES WP2.1. Conceptual framework for comparative multiple case study analysis

[22] 'Normative ambiguity points to the diverging ethical and normative assumptions in society.' WP2.1. Conceptual framework for comparative multiple case study analysis.

surrounding the term of CDSS, especially with regard to new systems.[23] It may thus be difficult to adequately categorize CDSS and thereby adequately examine their risks.

Ambiguity around the environment
Moreover, ambiguity exists to what extent an artificial system supports or replaces the decision-making of healthcare professionals in a CDSS (EPRS/STOA, 2019). This can be problematic when assessing to what extent an AI was responsible for a particular harm (was it, for example, the fault of the technology or the one that used it?) and how it thus can be prevented.

Ambiguity with regard to responsibility of harm is exacerbated when an algorithm is opaque, and due to the fact, that, especially in software development, components are sometimes 'blindly' borrowed or improved (from existing libraries for example) and treated as black boxes 'as long as it works'. The harm brought by an AI system could thus potentially be the result of a mistake from a previous developer (Mittelstadt et al. 2016) In the case of autonomous systems, the gap between a designer's control and the algorithm's behaviour can result in a situation where blame can be assigned to several moral agents simultaneously (accountability gap) (Ford and Price 2017).

Ambiguity around risk assessment
No clear consensus exists about how the possible risks surrounding AI should be characterized and ethically framed. Though the risks of AI were originally primarily framed in relation to safety, privacy and security, recent research has also pointed to the possible implications that AI (in healthcare) may have with regard to autonomy, distribution of power, human dignity, justice and control over technology (Kool et al. 2017), and the possibility as a society to guarantee certain human rights and civil liberties (EPRS/STOA 2019). The question how these values have to be weighed against each other makes the problem even more ambiguous.

Risk analysis of the use of CDSS is moreover surrounded by difficult ethical questions: what defines (human) responsibility? Can a machine really replace the essence of (good) 'care' and 'human contact'? How much of our privacy, intimacy, personal integrity, autonomy and power are we willing to trade for a healthier/longer life? These questions often do not have straightforward answers.

Normative ambiguity about risks is strengthened because the integration of AI in healthcare systems can be decisive for how the costs and benefits of these systems are distributed. The way such a system is developed and who gets a say in its conclusions brings forth ethical and political dilemmas. Moreover, different patients, healthcare professionals, managers, insurance companies and insurance payers, will have different perspectives on what counts as a risk, who will and should carry the burden of the risks (Wagner 2017).

Relevance of the precautionary principle to the case

In this section we analyse to what extent the precautionary principle may be of relevance to the risks of CDSS. We do this by checking to what extent the risks surrounding CDSS meet the requirements for application of the precautionary principle: 1. That the risks meet the threshold of damage. 2. That some form of scientific analysis has taken place 3. That there exists scientific uncertainty about the risks (Vos and Smedt, 2020).

Threshold of damage
It can be argued that, in terms of severity, the risks concerned with CDSS are comparable to the types of risks of other cases in which the precautionary principle has been applied.

[23] For instance, in the literature different terms are used and overlap exists of terms that refer to CDSS, like AI-Assisted Decision-making instead of CDSS, data driven CDSS / non-knowledge based CDSS.

In section 3.1 we showed that the risks surrounding CDSS are in principle directly proportional to the importance that decision-making by healthcare practitioners has in a society.

First of all, when a wrong decision is made on the basis of a CDSS this might thus amount to serious harm. Some argue that the precautionary principle can be understood as a modern restatement of the classical Hippocratic oath (Hanson 2018). In this respect, (a particular) use of AI in a CDSS may be forbidden because it may lead to avoidable/intentional harm. Other people even argue that a new Hippocratic oath is necessary for AI-developers (Etzioni 2018). However, it should be noted that 'human' decision making just as well can cause harm in healthcare and that CDSS may also prevent harm.

Secondly, the implementation of a CDSS can also be accompanied with public health risks. This is the case when a defective CDSS is implemented on a broad scale. If, for instance, multiple hospitals make use of the same system, it will have large scale effects when it doesn't function properly. Such risks may also spread when other technological systems (indirectly) make use of the data or the algorithms of the CDSS in question. A CDSS may also pose public health risks when its reasoning is used to support decision making that affects (large) groups, for instance in the case of support to decision making in epidemiology, population health and Learning Healthcare Systems.[24]

Thirdly, the precautionary principle has also been applied in the context of **human rights** and in particular circumstances the use of CDSS can be at odds with human rights. In section 3.1 we showed that CDSS are surrounded with a variety of data related risks, that may have implications for the right to access to healthcare, the right against discrimination, the right to respect for private and family life and the right to human dignity.

Fourthly, an implementation of CDSS can also result in severe power asymmetries and new dependencies of healthcare systems on outside actors. In some case this might lead to **irreversible consequences** that endanger the **sustainability** of the healthcare system. Precaution in this sense is prudent because the integration of AI in the decision-making of healthcare systems does have irreversible consequences on the moral principle of inter-generational equity. Legal scholar Joanna Mazur notes that there exist similarities between the nature of challenges faced in environmental law and data protection law. She argues that "if decision-making solutions were to pose a serious risk for public health or a high level of an unpredictability if applying these solutions in the policies referring to the protection of health, it might be possible to apply the precautionary principle as a legal measure to address the identified risks." (Mazur 2019).

Overall, it appears that the requirement of risks meeting the threshold of damage is met.

Some form of scientific analysis and scientific uncertainty

In section 3.2. we showed that some form of analysis has taken place with regard to the risks of CDSS. In section 3.3. we moreover described that the risks of CDSS are characterized by scientific uncertainty in a wide variety of ways. Both the technology of CDSS, the environment in which it is used (healthcare systems) and the difficulty of assessing the risks concerned, are characterized to some degree by ambiguity, complexity and uncertainty. Besides the relatively scattered status of the respective scientific disciplines concerned with these risks, the scientific uncertainty may thus be caused by a variety of aspects that are intrinsic to CDSS. All in all, it appears that some form of scientific analysis has taken place and that there is substantial uncertainty about the risks of CDSS.

Considerations contra invoking precautionary principle

Though the risks of CDSS potentially meet all of the criteria that makes it justifiable to invoke the precautionary principle, a variety of arguments can also be made to not invoke

[24] See for instance: Engler, A. (2020) A guide to healthy skepticism of artificial intelligence and coronavirus, The Brookings Institution. In which many of the hype around the use of AI for battling the corona virus are debunked and a variety of risks are addressed

the principle. First of all, it is important to note that healthcare is in itself a high-risk sector. Human decision making without CDSS just as well poses severe risks, and sometimes perhaps even more so. It is therefore crucial to assess the risks of a CDSS relatively to the risks that existing practices have, and also take into account that many CDSS may also prevent harm (see chapter 2, benefits).

Second of all, many of the reasons to invoke the precautionary principle in relation to CDSS are related to specific circumstances; the risks are highly context specific. Amongst others, they depend on the type of CDSS, their specific technical design, the situation in which they are used and the precautions that have been taken in the healthcare system. A CDSS that merely gives advice for harmless medical procedures does not seem to be in need of applying the precautionary principle. A CDSS that makes use of a good storage and authorization procedures around data has less need for precaution towards data risks. And, finally, as long as hospital keeps investing in the education of its personal, deskilling will also be less of an issue.

Relevance of the precautionary principle

Our analysis in the previous sections seems to indicate that the precautionary principle may be applicable to the use of CDSS, but only in specific circumstances. The principle may nevertheless be useful because it can point to appropriate regulatory and technical boundary setting for CDSS. In some scenario's, the seriousness of the risks clearly indicates the need for precaution (risks to public health, human rights), even when no scientific certainty about these risks has been established. Keeping these extreme situations and uncertainties in mind can inform decision making for taking the right precautionary measures; for instance, by limiting the medical procedures in which a CDSS can be used or the amount of human oversight that is necessary for important decisions.

Our analysis also shows that the risks of CDSS are in many cases difficult to define, both with regard to their specific outcome or harm, and with regard to their statistical probability. In these cases, the precautionary principle would be more suitable than, for example, the principle of prevention.

4 Risk governance and the precautionary principle

In this chapter we examine the risk governance that has taken place in the EU with regard to CDSS and the place the precautionary principle has had in it. The precautionary principle has not formally been applied in the EU by means of legislation or policies to the use of AI in healthcare, let alone the use of CDSS. Our analysis is therefore restricted to how in the EU was dealt with the risks of CDSS and to what extent precautionary thinking played a role.

We define risk governance as: "the totality of actors, rules, conventions, processes and mechanisms concerned with how relevant risk information is collected, analysed and communicated and management decisions are taken." (IRGC, 2018). Part of the risk governance are political and juridical dynamics (like legislation, regulation and policy initiatives) but also technological dynamics (like choices in the design of the technology) economic dynamics, (for instance markets with a high demand for safety and sustainability) and societal interactions or norms (for instance standards and practices among healthcare organizations for a safe use of CDSS).

In the first part of this chapter, we give an overview of the current legislation and regulation applicable to the EU that covers the risks of CDSS. The risk governance towards CDSS in the EU is in in a certain sense surrounded by a wide variety of legislation and regulation. Precaution towards the risks that may arise with CDSS is to some extent already covered by, for instance, regulation on medical devices and patient safety. These laws and regulations do not explicitly refer to CDSS, but nevertheless effectuate or can effectuate practices, norms and restrictions that impact how the risks of these systems are governed. For instance, while the GDPR does not explicitly refer to artificial intelligence or CDSS it does prescribe practices, norms and restrictions that are important for the data related risks of CDSS (for instance: data protection by default[25]).

In the second part of this chapter, we describe how EU risk governance has historically developed. We give an overview of the technological, economic, political and societal dynamics that played a role in how the risks of CDSS inside the EU have been dealt with. This analysis starts with the first precautionary warnings on AI in the research community (which largely happened outside the EU) and ends with forthcoming initiatives of the EU concerned with the use of AI in healthcare.

EU legislation and regulation

In this section we will give an overview of the existing regulation and legislation that covers some of the risks of CDSS in the EU.

It should be noted that (as described in sections 3.1-3.4) the risks of CDSS are complex, ambiguous and uncertain. This uncertainty, complexity and ambiguity applies to both the technological properties, the environment in which it is used as the assessment of risks concerned. As a consequence, different CDSS seem to fall in between different EU legislations and regulations. For instance, a data driven CDSS for instance generally has to refer to the GDPR to a larger extent than a knowledge based CDSS.

EU responsibility with regard to public health
It can be argued that the most extreme risks of CDSS are covered by a general responsibility of the EU to protect the public health of its citizens. However, the responsibilities towards the (public) health of the EU are limited.

The competence of the EU for public health has only been explicit inserted in the EU treaties since 1993, and is today laid down in Article 168[26] of the Treaty on the Functioning of the European Union.[27] It should also be noted that Article 168 entails a so-called supplementary competence which means that the EU can only supplement the actions undertaken by the Member States. Member States thus remain responsible for public health. Hence, EU action is required to respect 'the responsibilities of the Member States for the definition of their health policy and for the organisation and delivery of health services and medical care'.[28]

In the same fashion, the EU has to act and legislate consistently with the Charter of Fundamental Rights, but only to the extent that the EU has established competency over it.[29]

Indirect responsibility
Precautionary action on the EU-level towards CDSS could be inferred from more specific regulation. The risks of CDSS are covered by regulation on 1. safety of 'machines' in general. 2. medical products. 3. patient or consumer health and safety. 4. 'responsible' research and development. 5. Privacy.

[25] Article 25 of the EU GDPR
[26] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12008E168
[27] Article 168 states that: 'A high level of human health protection shall be ensured in the definition and implementation of all Union policies and activities.'
[28] Article 168 (7) TFEU.
[29] https://eur-lex.europa.eu/summary/chapter/human_rights.html?root_default=SUM_1_CODED%3D13

These regulations do not explicitly refer to CDSS and are therefore more 'technology neutral' with regard to risk governance.[30] We will now shortly describe to what extent these regulations indicate a precautionary approach towards CDSS.

Safety of 'machines' in general

There exist a variety of directives that are concerned with risks of the technology and materials that underlie (many) AI systems.[31] When 'precaution' is mentioned in these documents, it seems to be mainly concerned about the specific risks of the technology concerned (for instance that Electromagnetic equipment is accompanied by precautions that must be taken when the apparatus is assembled). In the Machinery Directive, moreover, no mentions seem to be made with regard to the precautionary principle or a precautionary approach.

Medical products and medical devices

EU regulation on medical products and medical devices consist of the Medical Devices Regulation, the Directive on Liability for Defective Products, The Directive On In Vitro Diagnostic Medical Devices, General Product Safety Directive, as well as laws of EudraLex; the collection of rules and regulations governing medicinal products in the European Union.[32] The fact that software and software integrated into devices have to be CE marked can in this sense be viewed as a precautionary measure.

In the case of an AI used in a CDSS in the form of a health app, it may be possible that the General Product Safety Directive is applicable. The General Product Safety Directive states that the precautionary principle can be used under certain conditions. Member States are expected to take measures 'in particular' where products 'could be dangerous' (Art. 8(1)(d)), are 'dangerous' (Art. 8(1)(e)) or where 'dangerous products [are] already on the market' (Art. 8(1)(f)).[33]

Patient or consumer health and safety

Precautionary action towards CDSS can also be inferred from the patient and consumer rights in the EU.[34] The precautionary principle is however not mentioned and the shared responsibilities seem limited. Illustrative in this respect is the following text of the Directive on the application of patients' rights in cross-border healthcare: 'As recognised by the Council (…) there is a set of operating principles that are shared by health systems throughout the Union. Those operating principles are necessary to ensure patients' trust in cross-border healthcare, which is necessary for achieving patient mobility as well as a high level of health protection. In the same statement, the Council recognised that the practical ways in which these values and principles become a reality vary significantly between Member States.'[35]

On the 12th of February this year, the European Parliament has however adopted a resolution in which it calls for a strong set of rights to protect consumers in the context of artificial intelligence and automated decision-making.[36]

[30] The GDPR, for instance, does not prescribe rules on data with regard to AI in particular, but speaks about data protection in general (independent of the type of technology that produces the data).

[31] such as the Low Voltage Directive, Electromagnetic Compatibility Directive and the Radio Equipment Directives.

[32] https://ec.europa.eu/health/documents/eudralex_nl

[33] See: WP 1, The effect of the precautionary principle since 2000

[34] For instance: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:088:0045:0065:EN:PDF

[35] DIRECTIVE 2011/24/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, (5), https://www.europarl.europa.eu/ftu/pdf/en/FTU_2.2.1.pdf

[36] https://www.europarl.europa.eu/news/en/press-room/20200206IPR72015/artificial-intelligence-meps-want-to-ensure-a-fair-and-safe-use-for-consumers

Responsible research and development
Fourthly, precaution can have a place in the R&D of CDSS and the regulatory framework of the EU around it. The notion of Responsible Research and Innovation (RRI) has been associated with the precautionary principle. René von Schomberg mentions the principle as one way to steer technological development in a societally desirable direction (Schomberg 2013). RRI is mentioned as a 'cross-cutting issue' in Horizon 2020, that will be promoted throughout Horizon 2020 objectives.[37]

Regulation on privacy
Finally, risks related to data accumulation are in a sense covered by legislation like the General Data Protection Regulation (GDPR). The GDPR protects citizens' fundamental right to data protection.[38] It is aimed, amongst others, to 'ensure a consistent and high level of protection of natural persons and to remove the obstacles to flows of personal data within the Union, the level of protection of the rights and freedoms of natural persons with regard to the processing of such data should be equivalent in all Member States.'[39]

The GDPR recognizes 'Data concerning health' as a special category of personal data.[40] It explicitly forbids taking decisions which produce legal or similarly significant effects for the individual solely in an automated way[41] and requires that the data subject should receive meaningful information on the logic involved in the process ('right to explanation').[42] This last provision has however not yet been enacted upon through jurisprudence and is questioned in academic literature (Wachter et al. 2017).

The precautionary principle or even the word 'precaution' are not mentioned in the GDPR. However, one could argue in some respects that similar types of reasoning are followed in the GDPR as in environmental legislation in which the precautionary principle is mentioned (Mazur 2019). The GDPR speaks of the implementation of the data protection 'by design and by default'.[43] Moreover, the requirement of a Data Protection Impact Assessment (DPIA) is defined in terms of 'Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons (….).'[44] The combination of the provisions inside the GDPR to 'ensure a high level of

[37] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

[38] GDPR, (1), 'The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her.'

[39] GDPR, (10).

[40] GDPR, Article 4 (15). 'Data concerning health' means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status;'

[41] 'The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' GDPR, Article 22 (1). This provision knows a few exceptions though: 'Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent.

[42] According to Articles 13(2)f, 14(2)g, and 15(1)h of the GDPR.

[43] GDPR, Article 25.

[44] GDPR, Article 35 (1) 'Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller

protection of natural persons', 'data protection by design and default', the requirement to perform a DPIA in the case of likely and high risks to the rights and freedoms of natural persons and the scientific uncertainty that is often ascribed to the effects of automated decision-making, seems to indicate to a possible invocation of the precautionary principle, so argues Joanna Mazur (Mazur 2019).

Other regulation
Other EU regulation that also deals with the possible risks of CDSS are rules on intellectual property, cyber security and trade regulation.

Moreover, some safety of CDSS may be covered by industry wide set standards, like from the International Organization for Standardization (ISO)[45] and the European standard (EN).[46] Interoperability of IT systems may for instance reduce some of the risks of CDSS because they make the CDSS more predictable and manageable.

There are also some individual companies that have developed their own ethical guidelines or norms for AI (Rathenau Institute 2019). Alphabet (Google), who recently has also entered the health market, has for instance described their principles on AI.[47] Philips has for instance launched 'Five guiding principles for responsible use of AI in healthcare and healthy living'.[48] It should be noted though that ethical guidelines, codes of conduct or other similar voluntary initiatives are not always very effective for risk governance (Del Castillo 2020).

Finally, there also exist standards, codes of conduct and best practices (like AI impact assessment) used by CDSS-developers, for instance Privacy and Ethics by design, that can reduce some of the risks (for more on this, see section: Effect of the precautionary principle on innovation pathways).

Legal cases
There exist a few legal cases on EU level that are relevant in relation to the risks of CDSS. One relevant court case is that European Court of Justice decided that software can be seen as a medical instrument, even when the software does not have a direct effect on the human body.[49] In this respect, precautionary measures in the context of regulation for medical devices could also be applicable for software-based CDSS.

Another court case that could be influential is the decision of the district court of the Hague, to shut down SyRI – An system Risk Indicator created by the Dutch Ministry of Social Affairs to identify people deemed to be at high risk of committing fraud – by citing the European human rights and data privacy laws. In this case the principles of proportionality and subsidiarity were also invoked, as it was shown that there are more privacy friendly alternatives that could have fulfilled the same aims.[50]

The European Court of Human Rights and the Court of Justice of the European Union have warned against the impact of surveillance activities from states on privacy rights (Van Est

shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.'

[45] The International Organization for Standardization (ISO) is an international standard-setting body. It is composed of representatives from various national standards organizations.

[46] The European standard (EN) is a standard for national standardization bodies of European member states.

[47] Google AI, Artificial Intelligence at Google: Our Principles.

[48] Philips (2020), Five guiding principles for responsible use of AI in healthcare and healthy living.

[49] ECJ, 7 December 2017 (Case C-329/16) 6

[50] https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865

et al., 2017). The European Court has moreover made multiple decisions on data protection.[51]

Other risk governance dynamics

In this section, we describe how EU risk governance has historically developed. We give an overview of the technological, economic, political and societal dynamics that played a role in how there has been dealt with the risks of CDSS inside the EU. This begins with the first developments and precautionary warnings on AI in the research community (which largely happened outside the EU) and end with forthcoming initiatives of the EU concerned with the use of AI in healthcare.

Early precautionary warnings inside the research community

Many of the main risks of CDSS that are currently discussed in the literature and among policy makers seem to have been voiced for a long time. Precautions were expressed about the fundamental limits of machine decision making, the danger that people would put too much trust in possibilities of the technology and the limits of what decision-making should be attributed to machines in the first place.

Already in the first centuries before Christ did people speculate on the thinking machine and the possible risks they might have (McCorduck 2004). More scientifically grounded assessments about risks of artificial intelligence however only emerged in the beginning of the twentieth century when the theoretical basis and components for constructing a thinking machine were increasingly thought to be in reach. Science fiction writer Isaac Asimov for instance introduced his Three Laws of Robotics in 1942. The first law - 'A robot may not injure a human being or, through inaction, allow a human being to come to harm.' – strongly resembles the medical no harm principle.[52]

The development of AI systems took off in the 1960's when AI research became heavily funded by the US Department of Defence and AI laboratories were established around the world. Criticism on the emerging field was expressed by Mortimer Taube in the book *Computers and Common Sense*. He argued that many AI research was only done on the premise of possible goals, without considering if such possibilities are enough of a justification to spend large amounts of money and time on it.

In the 1970's debates sparked up in the academic community about the (preferable) limits of AI.[53] Hubert Dreyfus argued in 1972 that human thinking could never be captured in formal rules, because it depends on unconscious processes (Dreyfus 1972). Researchers in artificial intelligence confused according to him the rule one is following to do something with the rule that can be used to describe someone doing something (McCorduck, 2004). Though a particular algorithm might perfectly describe someone's behaviour, this does not mean that it accounts for the internal deliberations that motivated the behaviour for example.

In the book Computer Power and Human Reason (1976), computer scientist Joseph Weizenbaum argued that robots should never be used to make important decisions, because they lack human qualities like compassion and wisdom. He also emphasized that machines would always lack the cultural and social background that play a role when humans make decisions. He stated that 'there are domains where computers ought not to intrude, whether or not it's feasible for them to do so.' And: 'Computers ought not be introduced where the effects can easily be seen to be irreversible and the side effects are

[51] https://ec.europa.eu/anti-fraud/sites/antifraud/files/caselaw_2001_2015_en.pdf

[52] This principle is thought to be part of the Hippocratic oath. It is often summarized with the phrase '"First do no harm'.

[53]'These debates were more academic (in the literal sense) than popular.' McCorduck (2004) Machines Who Think, 443.

not entirely foreseeable.' (McCorduck 2004) These contemplations retroactively show a strong similarity, with the principle of precaution.

Technological precautions taken with regard to the first CDSS

One of the first systems in which AI was used for support in medical decision making, called MYCIN, was developed in the early 70's. MYCIN consisted of approximately 600 rules that were used for making antibiotic treatment recommendations. These rules were based on facts about the patient and results of the antibiotic culture.

Considerations on the risks of such systems for the practice of healthcare professionals seem to be already part of the early development of such systems. Some developers for instance recognized the danger of a CDSS displaying too many messages ('alert fatigue') and the complexity of adjusting a specific expert system to the standards and practices of a specific healthcare system. For instance, the so-called CARE language was developed which allowed non-programmer, clinical experts to flexibly set the if-then-else logic of the alert according to their preferences (McCallie 2016).

Moreover, according to Kenneth W. Goodman the so-called 'Standard View' or standard apporach in computational diagnoses and the leading proponents of CDSS has always been caution. Randolph A. Miller, a 'key figure both in the scientific evolution of computational decision support and in scholarship on correlate ethical issues' has argued: "Limitations in man-machine interfaces, and, more importantly, in automated systems' ability to represent the broad variety of concepts relevant to clinical medicine, will prevent 'human assisted computer diagnosis' from being feasible for decades, if it is at all possible." (Goodman 2007).

The emergence of the first contours of EU risk governance on AI

EU risk governance of artificial intelligence also seems to emerge in the 1980's, in the wake of EU wide collaboration on research and the need for harmonization of IT standards, as well as the emergence of the first EU wide agencies (indirectly) concerned with the risks of technology.

In the 1980's the first EU wide research collaborations that focussed on new technology were started. Under the name ESPRIT (1983) a research programme was initiated to reverse the decline of European competiveness and to ensure global economic and political independence of European Communities in the face of the rise of the US and Japan in this field (Dorst et al. 2016). The programme had to result in better shared European protocols and standards in IT, for instance by financing large scale, long lasting, multi-country projects (a cooperative basis with industry, universities and governments of EC countries). Some of the projects were focused on advanced information processes which overall goal was to develop technological capabilities that underlie machine intelligence (Nilsson 2009).

From the 1980's onward, moreover, a variety of institutes and agencies were established that were concerned with or touched the governance of technology on a European level.[54] The 1980's also gave rise to a variety of ethical debates surrounding new technologies and the institutionalization of technology assessment around Europe (Schot and Rip 1997). In general, such debates and publications of EU wide agencies seemed first of all concerned with IT and digitization (and therefore only indirectly with AI) and their focus seems to be primarily on ethical, social and juridical aspects and not so much on risks.

In the 1990's the discipline of AI consisted of fragmented competing subfields focused on particular problems or approaches, often under different names (McCorduck 2004). This fragmentation may also partially explain why 'overarching' analyses concerned with risks

[54] Examples of such institutes are the Centre for European Policy Studies (1983), European Political Strategy Centre (1989), European Parliament's Panel for the Future of Science and Technology (1987), European Political Strategy Centre (1989), European Parliamentary Technology Assessment Network (1990), The European Institute of Innovation and Technology (2008) and the European Systemic Risk Board (2010)

of AI are hard to find; the development of an AI for a particular problem (like organizing a database) do not bring to mind substantial risks. In the early 2000's, a group of related technological developments promised revolutions with regard to AI capabilities in general[55], which thus again sparked a discussion about substantial and public risks.[56]

An early example of the explicit use of the principle in combination with AI is the report 'The Precautionary Principle in the Information Society Effects of Pervasive Computing on Health and Environment' (2003), by the TA-SWISS and STOA (Hilty et al. 2005). We have not been able to find, however, other analyses of AI, digitization and possible use of the precautionary principle on EU level.

EU risk governance in the wake of the digital single market

The risk governance towards CDSS in the EU significantly changed after the 2010's. In these years 'AI' and the use of AI in medical devices increasingly became an important concern in EU governance. Three developments contributed to the fact that AI and the risks of AI appeared at the forefront of EU policymaking.

First of all, new technological developments – the rise of big data, the growth of sophisticated machine learning techniques and cloud computing – created large expectations with regard to the (business) opportunities of AI.

Secondly, in the wake of the 'AI Revolution' a variety of intellectuals, politicians and societal organizations voiced concerns about the possible future societal risks of AI. Several ethics codes and principles for the development and use of Artificial Intelligence (AI) have subsequently emerged since 2017 from companies, partnerships between science, industry, and NGOs, and from politics and governance (Rathenau Institute 2019).

Thirdly, in the context of the aim to establish a digital single market, the development of AI became of a central economic and societal concern for the EU. Following the Lisbon Strategy, the Digital Agenda for Europe was initiated as one of the seven flagship initiatives of the Europe 2020 strategy.[57] In the context of this strategy, the Digital Single Market strategy sought 'to ensure better access for consumers and business to online goods and services across Europe, for example by removing barriers to cross-border e-commerce and access to online content while increasing consumer protection.'[58] In 2018 the Commission presented an AI strategy as part of the Digital Single Market Strategy. In its approach towards AI the Commission deals with technological, ethical, legal and socio-economic aspects 'to boost EU's research and industrial capacity and to put AI at the service of European citizens and economy.'[59]

Recent developments inside the EU

Many of the risks of CDSS are to some extent covered by existing EU regulation and legislation (see section on EU regulation and legislation). More recently, however, a variety of initiatives have emerged that relate specifically towards the risks of AI and the use of AI in healthcare, and therefore risks related to the use of CDSS. It is difficult to give a complete overview of all these initiatives and how they relate to each other. However, a few developments are worth mentioning.

First of all, efforts have been made to reduce ambiguity about the (legal/policy) definition of AI. The implementation of AI-specific legislation has possibly been complicated by the

[55] The rise of big data, the growth of sophisticated machine learning techniques and cloud computing

[56] In 2015, for instance, a collection of scientists and public intellectuals signed a open letter on which they pleaded caution with regard to the dangers of AI. They warn that 'systems must do what we want them to do.' https://futureoflife.org/ai-open-letter

[57] https://www.europarl.europa.eu/factsheets/en/sheet/64/digital-agenda-for-europe

[58] https://ec.europa.eu/eurostat/cache/infographs/ict/bloc-4.html

[59] https://ec.europa.eu/digital-single-market/en/artificial-intelligence

fact that, for a long time, no common understanding existed in the EU on what a robot or an AI system is. Recently, however, the European Parliament has defined what a 'smart robot' is[60] and the High-Level Expert Group on Artificial Intelligence (HLEG) has expanded on a definition of AI from the European Commission.[61]

Secondly, the EU has in collaboration with stakeholders and member states formulated ethical principles for AI and the contours of a specific European human centred approach. Since 2018 a wide variety of EU agencies have published general recommendations on AI (Rathenau Institute, 2019). In April 2018 moreover a declaration was signed by the EU members states in which they agreed to collaborate on the most important issues raised by AI.[62] In December 2018 the EC came with a Coordinated Plan on Artificial Intelligence,[63] in which it sketched out the intention of the EU becoming the world leader in the responsible development and application of AI. An EU high level expert group moreover developed ethical guidelines for AI,[64] an AI assessment list[65] and 'Policy and investment recommendations for trustworthy Artificial Intelligence'.[66] The Expert Group on Liability and New Technologies has moreover published a report about liability for artificial intelligence[67] and a forum - The European AI Alliance – was established for 'a broad and open discussion of all aspects of Artificial Intelligence development and its impacts.'[68]

Thirdly, the EU has implemented specific policies and stimulated collaboration with the specific aim to increase the use of AI in healthcare. There subsequently exist a certain technology push in the EU towards the implementation of CDSS. The European Commission strongly supports an enabling of the digital transformation of health and care in the Digital Single Market.[69] The Commission argues that only by fundamentally rethinking the EU health and care systems, it can be ensured that they remain fit-for-purpose. The Commission mentions ageing, multimorbidity, a growing threat from infectious diseases due to increased resistance to antibiotics and new or re-emerging pathogens, health workforce shortages, and the rising burden of preventable noncommunicable diseases caused by risk factors such as tobacco, alcohol, and obesity, as some of the main challenges that may need 'digital solutions'.[70] The Commission argues that market fragmentation and lack of interoperability across health systems currently stand in a way of an integrated approach for the EU. Such an approach is difficult because the organisation and delivery of healthcare is the responsibility of the Member States, and in some Members States the financing and provision of healthcare is even the responsibility of regional authorities.

Fourthly, – more recently – the EU has developed a risk-based approach toward some forms of AI. In the White Paper on AI of 2020, the European Commission proposes specific assessment requirements for 'high-risk' AI applications, depending on the sector in which it is deployed and the manner in which it is deployed. Healthcare is mentioned as a high-

[60] https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html
[61] https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines
[62] https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/eu-declaration-cooperation-artificial-intelligence
[63] https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX:52018DC0795
[64] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[65] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/2
[66] https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence
[67] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=63199
[68] https://ec.europa.eu/digital-single-market/en/european-ai-alliance
[69] https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering
[70] https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering

risk sector, but not all applications are high-risk: ‘For example, whilst healthcare generally may well be a relevant sector, a flaw in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. The assessment of the level of risk of a given use could be based on the impact on the affected parties.’ (European Commission, 2020).

Besides applications that fall under the two criteria, some exceptions could also be considered high-risk, like ‘the use of AI applications for the purposes of remote biometric identification and other intrusive surveillance technologies.’ The European Commission in this regard mentions pre-marketing conformity assessment requirements, requirements on training data, requirements on record-keeping and data sets, requirements on human oversight, transparency, accuracy and human oversight, monitoring and ex-post controls. The Commission also proposes the establishment of a Code of Conduct for processing personal data in the health sector.

Lastly, the EU has mentioned a variety of forthcoming regulation and revisions of existing legislation with, amongst others, the aim to tackle AI specific risks. The EU does not yet have specific legislation on robotics or AI, but the European Commission is expected to implement (binding) regulatory and policy initiatives in the following years (Molyneux et al. 2017).

The European Parliament has moreover requested to examine legal questions in connection to the development and use of robotics and artificial intelligence foreseeable in the next 10 to 15 years. The Commission has subsequently launched an evaluation of the Directive on Liability for Defective Products, Expert Group on Liability and New Technologies and the Machinery Directive.

In a Resolution on 12 February of this year, the European Parliament has called for a strong set of rights to protect consumers in the context of artificial intelligence and automated decision-making.[71] The Parliament argued that automated decision-making (ADM) technologies should only make use of unbiased data sets and explainable and unbiased algorithms, with review structures set up to remedy mistakes and the possibility of consumers to redress automated decisions. Those systems should only use high-quality and unbiased data sets and “explainable and unbiased algorithms”, states the resolution. Review structures should be set up to remedy possible mistakes in automated decisions. It should also be possible for consumers to seek redress for automated decisions that are final and permanent: ‘“Humans must always be ultimately responsible for, and able to overrule, decisions that are taken in the context of professional services such as the medical, legal and accounting professions, and for the banking sector.”

5 The precautionary principle and its future

Reflection on the precautionary principle in the literature

The precautionary principle has not yet been explicitly applied in EU legislation or regulation, in relation to CDSS. It is no surprise, therefore, that explicit reflection on the application of the precautionary principle to CDSS has been limited.

Though criticism on the application of the precautionary principle specifically in relation to CDSS or on the use of AI in healthcare seems scarce, comments have been made on the use of the principle in relation to AI in general. Opponents of the precautionary principle

[71] https://www.europarl.europa.eu/news/en/press-room/20200206IPR72015/artificial-intelligence-meps-want-to-ensure-a-fair-and-safe-use-for-consumers

Daniel Castro[72] and Michael McLaughlin[73] argue that it undermines progress in artificial intelligence (Castro and McLaughlin 2019). They refer to a very strong interpretation of the principle: 'The precautionary principle is the idea that if a technological innovation may carry a risk of harming the public or the environment, then those proposing the technology should bear the burden of proving it will not. If they cannot, governments should limit the use of the new technology until proven safe. Those who support the precautionary principle, which call for government intervention even when there is no clear evidence of tangible and imminent threats of harm, adhere to the cliché it is "better to be safe than sorry."' (Castro and McLaughlin 2019).

Castro and McLaughlin state that the application of the precautionary principle in relation to AI leads to slower and more expensive AI development, less innovation, lower-quality AI, less AI adoption, less economic growth, fewer options for consumers, higher prices, inferior consumer experiences, fewer positive social Impacts and Reduced Economic Competitiveness and National Security.

There are also proponents of the precautionary principle. The European Trade Union Institute (ETUI) has called for the precautionary principle and human rights in their Foresight Brief about the need for regulation for workers in the context of AI (Del Castillo 2020). ETUI argues, amongst others, that 'the precautionary principle is an essential principle that must be at the heart of technological development. It can sustain such development, give direction to innovation and, in the case of AI, help to (1) build a governance based on social dialogue and which involves relevant societal actors; (2) provide a framework conducive to the explicability and accountability of algorithmic decision-making; (3) contribute to ensuring that technological innovations are safe for society.' Moreover, the innovation principle is described as '(…) a concept which was invented in 2013 by various CEOs as a lobbying/deregulatory tool and which does not have a legal basis. It is not found in EU treaties, secondary legislation, case law or the national constitutional traditions of any Member State.'

Effect of the precautionary principle on innovation pathways

The precautionary principle has not explicitly been applied to the use of CDSS, but precautionary thinking has in different examples had an effect on the development of these systems. First, we look at the general (geopolitical) background that influences the innovation pathways of AI development, and thereby the innovation pathways of CDSS. In the second part we examine some of the ways in which other choices are or can be made on the basis of precaution in the development of CDSS.

General (geopolitical) background

The development of AI worldwide is often portrayed as a race, whereby a leading position is deemed essential for national security (Hunter et al. 2018) and/or economic security (McKinsey Global Institute 2018). Because the AI also poses substantial risks – and to attain mitigate such risks as well as assure legal and economic certainty – this has also led to a 'race to AI regulation'. Good regulation could also effectuate a regulatory 'first mover advantage' (Smuha 2019). Besides the EU, also Japan, Canada, Dubai, China, Singapore, the US and Australia have published ethics guidelines for AI (Smuha 2019). Moreover, besides risk-regulation, there exists competition in leading the technological standard-setting processes, which can also have ethical consequences (Beatie, 2019).

Moreover, the potential impact of AI has been subject of wide speculation, from those that characterize it as a fundamental tool for defence or who see it as an inevitable step towards

[72] Daniel Castro is vice president at the Information Technology and Innovation Foundation (ITIF) and director of ITIF's Center for Data Innovation. - https://itif.org/person/daniel-castro

[73] Michael McLaughlin is a research analyst at the Information Technology and Innovation Foundation. - https://itif.org/person/michael-mclaughlin

singularity (Creighton 2018), to those that warn for its possibilities of totalitarian control (Helbing et al. 2018). AI is in this sense a 'controversial' technology, a fact that may slow down a steady uptake of the technology.

Another factor that influences the innovation pathways of CDSS is that many of the AI applications that are currently featured in medical literature, are not easily executable at in clinical practice: 'A complex web of ingrained political and economic factors as well as the proximal influence of medical practice norms and commercial interests determine the way healthcare is delivered.' (Panch et al. 2019). Secondly, in many healthcare organizations the necessary data infrastructure to collect data and train an AI, and test for possible biases, is lacking. Besides regulatory uncertainty, a variety of established customs and conservative views and interests may play a role in the innovation path (see also environmental variability section 3.1).

Nevertheless, with these dynamics in the background, different countries follow different strategies with regard to AI. These national strategies set, as it were, the stage within which R&D on AI and CDSS is acted out. In the case of the EU, the background is characterized by the need for harmonization of the regulatory framework of the member states in service of the digital internal market, the conviction that AI can help to solve some of the world's biggest challenges and its human-centric approach to AI (see 3.3.1). In other regions, other political and economic factors play a (more decisive) role. We will shortly look at the strategies of China and the United States.

China

In 2017 the State Council of China released the 'New Generation Artificial Intelligence Development Plan.'[74] China strives to be the leading AI superpower in 2030, largely through state funding and considers the development of AI a national priority. It is part of the state-driven industrial plan 'Made in China 2025'. In May 2019, a multistakeholder coalition consisting of Chinese universities, the Institute of Automation and Institute of Computing Technology in Chinese Academy of Sciences, and firms like Baidu, Alibaba and Tencent, developed the Beijing AI Principles. They are 'proposed as an initiative for the research, development, use, governance and long-term planning of AI, calling for its healthy development to support the construction of a human community with a shared future, and the realization of beneficial AI for humankind and nature.'[75]

The United States

The United States also considers worldwide leadership in AI as a national priority. On February 11 2019 the American AI Initiative was launched. In it is stressed that 'the Federal Government plays an important role not only in facilitating AI R&D, but also in promoting trust, training people for a changing workforce, and protecting national interests, security, and values.' The initiative is guided by five principles: 1. Driving technological breakthroughs, 2. Driving the development of appropriate technical standards, 3. Training workers with the skills to develop and apply AI technologies, 4. Protecting American values including civil liberties and privacy and fostering public trust and confidence in AI technologies, 5. Protecting US technological advantage in AI, while promoting an international environment that supports innovation.

Choices in the design of CDSS

Besides the geopolitical background, the innovation pathways of CDSS are mainly dependent on the specific choices that are made by the developers of these systems. These choices are often determinative for the risks that CDSS pose. The design choices that are made with regard to clinical decision support systems depend naturally on the specific

[74] Future of Life Institute (visited 9 April 2020), AI Policy China, https://futureoflife.org/ai-policy-china/.

[75] Bejing AI Principles, https://www.baai.ac.cn/blog/beijing-ai-principles.

function it serves (for instance the extent to which malfunction would lead to harm). Nevertheless, a few general differences in design choices can be observed.[76]

First of all, choices are made with regard to the data; which data is used, how and where it is stored, shared and processed and who has access to it. The patient data that a CDSS makes use of can for instance be stored at a decentralized location and when it is centrally stored it can be anonymized. Data collection may furthermore be checked for biases and if its algorithms are up to date with contemporary medical knowledge and reasoning.

Secondly, differences exist in how the CDSS comes to conclusions; the technique that is used to reason. This may differ from a machine learning approach that is purely based on data, or a structure that is based on medical knowledge trees (knowledge based vs data based). And in the case of machine learning, a distinction is made between supervised or unsupervised machine learning. In the development of some CDSS, it is monitored by the developers whether the conclusions of the CDSS are in agreement with the conclusions made by real doctors and if their use indeed lead to better results. In various instances CDSS have underwent clinical trials.[77] Accountability of the decision-making can moreover be improved by making use of explainable AI (XAI).

Thirdly, in a variety of ways developers have thought about the layout of CDSS; about how they influence medical decision making in a good way; promoting reflection and calmness. This too may ensure that healthcare professionals stay in control. An example of this are choices in how and how often the alerts are showed to a practitioner (for instance to reduce alert fatigue or stress). The way a message may be presented (coercive, interactive etc) may also be taken into consideration.

Fourthly, choices in the programming language, the software and the hardware that is used may be decisive for the accessibility and flexibility of a CDSS. Some CDSS make use of flexible coding so that healthcare professionals can easily adjust it to their preferences. Interoperability may increase the availability of support and information about particular systems, but it can also lead to the situation where the market is in control of a few companies. This may lead to undesirable dependencies.

Moreover, many of the uncertainties caused by environmental variability of healthcare systems can be reduced by involving stakeholders in the design process of a CDSS. This can ensure that a CDSS is better attuned to the work flow, expectations and requirements of healthcare professionals. This makes the use of the CDSS more predictable for its users, which also diminishes risks. In some instances this might for example prevent a misdiagnosis.

In this sense, through precautionary approaches or the application of the precautionary principle there are a variety of possibilities in which the innovation pathway of a CDSS can be steered into a more ‘risk-free’ direction.

Innovation principle

The innovation principle has not been applied in relation to artificial intelligence, let alone the use of CDSS. As of yet, no policies, laws and regulation on AI can be found that make use of the principle.

[76] For instance: Zikos, D. and DeLellis, N. (2018) CDSS-RM: A clinical decision support system reference model. BMC Medical Research Methodology. 18. 10.1186/s12874-018-0587-6. See also: ‘References’ Medicine / Health information technology.

[77] See for instance: Jia P et al. (2016) The Effects of Clinical Decision Support Systems on Medication Safety: An Overview. PloS ONE 11(12).

There has however been some, but not many, discussions about the innovation principle in relation to AI. In an article, Daniel Castro[78] and Michael McLaughlin[79] advise that the innovation principle instead of the precautionary principle should be applied by policy makers when AI is concerned. They juxtapose the innovation principle to the precautionary principle: 'While some people advocate for an almost completely hands-off approach to regulating new technologies, those who recognize that there is a legitimate role for government take two distinct approaches toward action: the precautionary principle and the innovation principle.' (Castro and McLaughlin 2019).

They relate the innovation principle to the conviction that '(…) because the overwhelming majority of technological innovations benefit society and pose modest and not irreversible risks, government's role should be to pave the way for widespread innovation while building guardrails, where necessary, to limit harms.' Moreover, they emphasize that the innovation principle – which they define as the principle that '(..) the vast majority of new innovations are beneficial and pose little risk, so government should encourage them' - recognizes 'that market forces, tort law, existing laws and regulations, or light-touch targeted interventions can usually manage the risks new technologies pose.' And that it advocates case-by-case regulation and that, in cases where regulation is needed, it 'stresses the importance of designing regulatory interventions and structuring regulatory enforcement in ways that minimize the harm to innovation, while still achieving the regulatory goals.' Finally, the principle focusses, according to them, 'on ensuring that penalties punish bad actors who cause harm than creating regulations that limit beneficial and benign uses.'

In relation to AI, embracing the innovation principle would, they argue, allow society to experience the benefits of AI 'while adopting the right, limited regulatory frameworks that enable innovation while limiting harms.'

The European Commission has also connected the innovation principle with AI in a communication on AI in 2018. In a footnote the EC writes: 'For any new regulatory proposals that shall be needed to address emerging issues resulting from AI and related technologies, the Commission applies the Innovation Principle, a set of tools and guidelines that was developed to ensure that all Commission initiatives are innovation friendly.' [80]

The European Commission also mentions the innovation principle on its website as a tool 'to help achieve EU policy objectives by ensuring that legislation is designed in a way that creates the best possible conditions for innovation to flourish.'[81] The Commission states that 'the possible effects of emerging technologies on EU rules should be scrutinized early in the legislative process as part of the Innovation Principle.' AI is mentioned as an example of such an emerging technology. Notably, in this formulation the innovation principle does seem to be in opposition with the precautionary principle.

The Centre for European Policy Studies also mentions AI in its 'study supporting the interim evaluation of the innovation principle'. (Renda and Simonelli 2019). They write that the application of the innovation principle 'would intuitively need to go hand-in-hand with reflecting on and developing experimental regulation' in areas such as artificial intelligence (Renda and Simonelli 2019).

[78] 'Daniel Castro is vice president at the Information Technology and Innovation Foundation (ITIF) and director of ITIF's Center for Data Innovation.' - https://itif.org/person/daniel-castro

[79] 'Michael McLaughlin is a research analyst at the Information Technology and Innovation Foundation.' - https://itif.org/person/michael-mclaughlin

[80] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

[81] https://ec.europa.eu/info/research-and-innovation/law-and-regulations/innovation-friendly-legislation_en

6 Synthesis

The aim of this case study was to better understand the complexities and controversies of applying the precautionary principle to the use of AI in healthcare. We focused on clinical decision support systems (CDSS), because the risks surrounding these systems are exemplary for the general complexities and problems that surround the use of AI in healthcare.

Clinical Decision Support Systems (CDSS) are, in a broad sense (as the name implies) systems that support the decision making of healthcare practitioners. These systems for example provide alerts or reminders, highlight guidelines during care, provide suggested course of action and identify drug-drug interaction. Proponents argue that CDSS provide faster, more accurate decision making with less costs and human errors. In some cases, CDSS might even make new decision-making (on the basis of Big data) possible that could improve the overall efficiency and effectivity in healthcare. It should be noted that there exists considerable uncertainty with regard to many CDSS applications about these possible benefits, especially with regard to long term benefits and the extra costs of, for instance, maintenance.

CDSS replace, augment or supplement decision making processes in healthcare. The use of CDSS is thus accompanied by risks because such decisions can have large impact. A wrong decision in the domain of healthcare can potentially have severe effects on individual health, human rights and – if a CDSS is implemented on a broad scale or if it supports decisions on groups – public health. Although human decision making in healthcare is also accompanied by such risks, CDSS also pose *new* risks to the extent that they transform *how* such decisions are made: their decisions are exclusively based on data, they are based on machine reasoning (and therefore lack human elements), they imply a delegation of control from the patient or healthcare practitioner to a machine, and their use is accompanied by a new division of labour in the healthcare domain.

The risks of CDSS

Taken to its extreme, and if no precautions are taken, the transformation of decision making in a healthcare system by CDSS can have severe consequences. Overconfidence about the capabilities of AI in combination with biased or defective datasets/algorithms, for instance with regard to gender/sex bias in medical data, can cause unnecessary deaths or disease. Moreover, especially data driven CDSS are accompanied by a variety of data related risks, like infringements on the right to privacy and the involvement of unqualified actors into the norm setting of medicine. And, finally, health data that has fallen into the wrong hands can be used against people, like blackmailing, and can be used as a tool to predict and manipulate future behaviour. This primarily has consequences for the distribution of power, equal access to public benefits and the right not to be discriminated.

The use of CDSS can also imply a delegation of control from the healthcare practitioner and the patient. This can endanger the autonomy of these actors and can also lead to deskilling and accountability gaps. Taking away the human element in the decision making could infringe on the right to healthcare to the extent that care necessitates a person that ‘cares for’ or is ‘involved’ your suffering. Finally, the replacement of decision making in healthcare with CDSS, tends to be accompanied by a new division of labour. Other actors, like IT companies and data collection agencies, acquire a (more important) place in the domain of healthcare. This can bring forth new dependencies and therefore new risks, for instance rising costs due to locked ins in suppliers and maintenance, which can have consequences to the affordability of and access to healthcare.

Complexity, ambiguity and uncertainty around the risks of CDSS

The risks mentioned above are all characterized by a high degree of uncertainty: both with regard to their precise effects and with regard to their probability. First of all, this uncertainty is highly dependent on the specific technological properties of a CDSS. It can display complex and uncertain behaviour, especially when it makes use of unsupervised machine learning, uncertainty to the extent that small changes in its core code have significant effects and epistemic uncertainty to the extent that its code and connections to other systems are inaccessible and not understandable. Ambiguity may be an issue with regard to understanding 'why' a CDSS has made a particular suggestion.

Secondly, the use of CDSS is characterized by uncertain risks due to the nature of the environment in which it is implemented. Healthcare systems can be complex, unpredictable systems and the role a CDSS fulfils for each of these actors can be ambiguous. For a safe implementation of a CDSS in a healthcare system, for instance a hospital, it has to be attuned to a system that exists of many interacting and unpredictable elements. The CDSS for instance has to be in line with the (changing) expectations, protocols and existing norms and standards of healthcare professionals. A CDSS for instance has to be readable, understandable and helpful in the context of the daily tasks of a doctor, the specific needs of a patient and the oversight of a manager and/or a privacy officer. Some CDSS moreover have to mediate between different aims, standards, inputs and multiple different sets of data or other IT systems. The interaction of CDSS with other systems and actors can lead to feedback loops, especially when it is data driven: it can change according to for instance, the patients that are included in its data, data about the decisions that a doctor has made or updates of its algorithms. This can make them unpredictable. A CDSS moreover has to be attuned to the inherent uncertainty that exists in healthcare when it comes to complex, ethically complex or unknown medical problems.

A third cause for the uncertainty around the risks of CDSS is the variability in the nature of the risks, which makes them difficult to assess. To the extent that 'good' or safe decision making in healthcare consists of many elements, so the risks can be a consequence of multiple elements. A good decision is for instance transparent, explainable, accountable, supported by representative data, sufficient reflection, respect for privacy, autonomy and dignity of the patient. A safe use of a CDSS needs to take into account each of these elements, but these elements are ambiguous.

Scientific uncertainty

The fact that both the technology of CDSS, the environment in which they are used and the assessment of risks are characterized by uncertainty has consequences for the possibility of analysing them scientifically. First of all, because of this the scientific analysis of them is scattered over a wide variety of scientific disciplines. An adequate analysis of the risks of a CDSS has to make use of knowledge from, amongst others, the field of AI, medical professionals, legal scholars, and medical ethicists.

Moreover, it seems to be a challenge to develop uniform criteria to assess the risks of CDSS because each implementation of a CDSS is somewhat unique with regard to the technical characteristics of the system, the environment in which it is used and the precautions that are already taken in this environment. Finally, new developments of CDSS happen fast and many data driven CDSS are relatively new. All of this seems to contribute that the fact that there does not seem to be a clear scientific consensus or certainty about the risks of CDSS or how they should be assessed.

Risk governance of CDSS

It is difficult to make firm conclusions about the risk governance of CDSS, partially because this seems to be still in process. A few things can be discerned however that are notable in the context of the complexities and controversies in the case.

First of all, precaution towards the limits and risks of CDSS was already voiced early on by a variety of researchers in the field of AI. Many of their concerns – for instance with regard to control over AI and the limits of machine reasoning – overlap with the concerns that are still at issue in EU policy debates.

Secondly, precautionary thinking about the specific design of CDSS also seems to have been present early on. Key figures in computational decision argued for precaution and adjustments were developed with regard to programming languages and notification systems.

Thirdly, EU risk governance around CDSS seems to have emerged in the context of a strong economic incentives. The contours of this emerged in the 1980's when the EU began collaborations on AI research to compete with the rise of Japan and the US. In the 2010's AI increasingly became of a central concern in the wake of the establishment of the digital single market.

Fourthly, we showed that the risks of CDSS have been embedded in a complex web of EU legislation. They may be (partially) covered by legislation on safety of machines, medical products, patient or consumer health and safety, regulation on 'responsible' research and development, privacy, intellectual property, cyber security and trade regulation, as well as a few legal cases.

To reduce complexity and legal uncertainty, the European Commission has recently undertaken a variety of initiatives that are more specifically aimed at AI and the risks of AI (in healthcare). In these initiatives the EU distinguishes itself from other geographical areas through cooperation with ethicists, AI researchers, businesses, consumer organizations and other stakeholders and close coordination between the member states. Multiple existing EU legislations are under review to align them with the specificities of AI and multiple ethical guidelines have been published. Notably, these initiatives first of all seem to have an ethical focus. Only in the recent White Paper on AI, published in February this year, did the EU explicitly adopt a risk-based approach in which the use of AI in healthcare was defined as 'high-risk'. As of writing, this paper is up for public review.

The relevance of the precautionary principle and the innovation principle

The precautionary principle seems to be potentially applicable to CDSS, but only on a strict case by case basis: for instance depending on the type of CDSS (especially data driven CDSS), the nature of the decision (for instance: when public health or communicable diseases are concerned), the type of data (for example: biometric data), how it contributes to the decision-making (for example: automatic, in absence of any human reasoning) and the place of the CDSS within a particular health environment (for example: when it is intertwined with a wide variety of processes in a hospital).

In extreme cases the risks of implementing a CDSS meet the criteria of the threshold of damage (public health and human rights). Moreover, scientifically grounded analysis has been done on these risks, but there remains significant scientific uncertainty about both the precise nature of the possible harmful outcomes and the probability of these outcomes are uncertain (See also conclusion).

The innovation principle does not seem to be particularly relevant in this case. Careful considerations about the uncertainties and requirements of CDSS in the vulnerable domain of healthcare, logically seem to have the upper hand over the benefits of innovation in terms of jobs and economic growth or the health benefits that CDSS may offer in the long run. Especially because many of the risks surrounding CDSS are about the question if the automation of decision making is desirable and beneficial in the first place. However, this too should be examined on a case by case basis.

What can we learn from this case in relation to other RECIPES cases?

An important difference between this case study and the other RECIPES cases is that this case is concerned with if the precautionary principle *might* be applicable, why it has not been applied and to what extent other risk governance has been undertaken. Answers to these questions should primarily follow from the cross-case comparison, but on the basis of this case there are a few possible answers:

1. The precautionary principle has historically mostly been applied to environmental risks (and more recently public health). Though risks regarding CDSS can also be quite severe, they do not relate to the environment. There are, as Joanna Mazur notes (Mazur 2019), nevertheless similarities between the nature of challenges faced in the area of the data protection laws and environmental laws.
2. Many of the most serious risks of CDSS are related to the violation of human rights, like autonomy, equal access to healthcare and privacy. The precautionary principle has been acknowledged by the European Court of Human Rights (EHRM) in relation to human rights.[82] It should be noted though that the application of the principle in relation to human rights does not seems to be an established practice.
3. Problematic applications of CDSS are relatively recent. Only since the 2000's, in the wake of the AI revolution, have questions around (data driven) CDSS become urgent (for the EU). Many of the risks of CDSS are new 'types' of risks. While risks related to public health and the environment have been publicly discussed and institutionalized for a long time, questions concerning autonomy and power asymmetry in relation to big data are newer. Moreover, they are often primary discussed in terms of ethical or philosophical questions and/or difficult to formalize in risk assessment standards. The precautionary principle could be of relevance to these type of risks because a growing body of research indicates that these risks can also be systemic, irreversible and that they are connected with the violation of human rights.
4. In most other RECIPES cases the precautionary principle is applicable because the risks have to do with biological systems. The implications of, for instance, GMO's, are considered to be severe, disruptive and irreversible because they can influence the dynamics of ecological systems. Because these systems are alive, changing and dynamic, such risks are difficult to predict and control. In contrast, CDSS, and AI systems in general, are (generally) geographically closed off systems. It should however be noted that a disruption of a healthcare system by a CDSS can also have additional effects on societies as a whole. If, for instance, a hospital can no longer provide care due to disruptive effects of an AI this may do severe physical, emotional and psychological harm to those who depend on the services of the hospital. This in turn may strain the resilience of the society as a whole.
5. Related to point 4; while the other cases are primarily concerned with risks that arise due to the interaction of humans with the environment (and the long-term effects this may have), this case is primarily about risks that ultimately come down to 'interaction' between humans. CDSS are made by humans, for humans, used by humans, on humans. Subsequently, many of the risks around the implementation of CDSS are, more than in the other cases perhaps, about political and socioeconomical dynamics (and therefore human rights). It is very much about how to balance the different interests and power relations that can be at play in the decision making in healthcare. It is about the humanity, autonomy and privacy that is deemed necessary. The rights of the patient and healthcare practitioner, balanced with, for instance, efficiency, better standards of living, longer life expectancy and economic growth.

[82] Tătar EHRM 27 januari 2009, ECLI:CE:ECHR:2009:0127JUD006702101 (Tătar/Roemenië).

7 Conclusion

Our analysis indicates that the precautionary principle is in theory applicable to clinical decision support systems (CDSS), but only in particular cases. A careless implementation of CDSS into healthcare systems can, especially in the case of decisions that affect groups and for systems that are implemented on a large scale, bring significant harm, both with regard to individual health, public health and human rights. The criteria of scientific analysis of these risks, on which it should be decided whether the precautionary principle is relevant for risks associated with CDSS, also seem to be met. Knowledge and empirical findings on the risks of CDSS or similar AI systems, insights on the vulnerability of healthcare systems and health data, as well as examples of problematic usage of AI in decision making processes in healthcare and other sectors, do warrant, in our opinion, invoking the precautionary principle.

Moreover, as our risk analysis shows, the risks of CDSS are in many cases difficult to define, both with regard to their specific outcome or harm, and with regard to their statistical probability. In these cases, the precautionary principle would be more suitable than, for example, the principle of prevention.

It should however also be noted that many of the reasons to invoke the precautionary principle in relation to CDSS are related to very specific circumstances; the risks are highly context specific. Amongst others, they depend on the type of CDSS, its specific technical design, the situation in which it is used and the precautions that already have been taken. For instance, a CDSS that merely gives advice for harmless medical procedures does not seem to be in need of applying the precautionary principle. A CDSS that makes use of a good storage and authorization procedures around data has less need for precaution towards data risks. And, finally, as long as a hospital keeps investing in the education of its personal, deskilling will probably also not be an issue.

In this regard, criteria can be developed by policy makers that point to circumstances in which the precautionary principle is especially relevant in relation to the implementation of a CDSS. This case would suggest criteria based on the type of CDSS (especially data driven CDSS), the nature of the decision (for instance: when public health or communicable diseases are concerned), the type of data (for example: biometric data), how it contributes to the decision-making (for example: automatic, in absence of any human reasoning) and the place of the CDSS within a particular health environment (for example: when it is intertwined with a wide variety of processes in a hospital).

The precautionary principle can be useful in multiple ways. The principle first of all can be instrumental for delineating the limits of the implementation of CDSS. Policy makers, healthcare professionals and companies could ask themselves what the minimal requirements for a safe and good decision-making process in healthcare are. Which decisions should always be taken by a human or in deliberation with the patient? What are the minimal requirements of a decision to make it sufficiently accountable, transparent, evidence based and uninfluenced by non-medical considerations or interests? Which type of health data or combinations of data should never be used outside of medical practices?

Secondly, the precautionary principle can stimulate reflexivity and awareness of the many uncertainties around the implementations of CDSS. We showed that the risks of CDSS are characterized by many uncertainties, because of the nature of the technology, the properties of healthcare systems and the types of risks that are concerned. To this extent the precautionary principle may encourage anticipation, cocreation and incremental innovation of CDSS.

8 References

Scientific literature

AI research field and computer science

- Dreyfus, H. L. (1972) What Computers Can't Do: A Critique of Artificial Reason, Harper & Row.
- Weizenbaum, J. (1976), Computer Power and Human Reason: From judgment to calculation. W. H. Freeman & Co.
- Russel, J. R. and Norvig, P. (2010), Artificial Intelligence: A Modern Approach, Pearson Education Limited.
- Nilsson, N. J. (2019) The Quest for Artificial Intelligence: A History of Ideas and Achievements, Stanford University, p. 355 – 357.
- McCorduck, P. (2004) Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence

(Bio)-ethics and philosophy

- Bali, J., Garg, R., & Bali, R. T. (2019). Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? Indian journal of ophthalmology, 67(1), 3–6. https://doi.org/10.4103/ijo.IJO_1292_18
- Nuffield Council on Bioethics (2018), Bioethics briefing note: Artificial Intelligence (AI) in healthcare and research, https://www.nuffieldbioethics.org/news/big-ethical-questions-artificial-intelligence-ai-healthcare.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society. https://doi.org/10.1177/2053951716679679
- Goodman K.W. (2007) Ethical and Legal Issues in Decision Support. In: Berner E.S. (eds) Clinical Decision Support Systems. Health Informatics. Springer, New York, NY.

Science & Technology Studies and Technology Assessment

- Kool, L., J. Timmer, L. Royakkers and R. van Est, Urgent Upgrade - Protect public values in our digitized society. The Hague, Rathenau Institute 2017.
- Niezen, M.G.H., Edelenbosch, R., Van Bodegom, L. and Verhoef, P. (2019). Health at the centre – Responsible data sharing in the digital society. The Hague: Rathenau Institute.
- Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool (2017), Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality – Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE), The Hague: Rathenau Institute.
- Schot, J. and Rip, A. (1997), The Past and Future of Constructive Technology Assessment, Technological Forecasting and Social Change Volume 54, Issues 2–3, 251-268. https://doi.org/10.1016/S0040-1625(96)00180-1.
- Hilty, L. et al. (2005). The Precaution Principle in the Information Society. Effects of Pervasive Computing on Health and Environment. Report of the Centre for Technology Assessment.

- Rathenau Institute (2019), Overview of ethics codes and principles for AI, https://www.rathenau.nl/en/digital-society/overview-ethics-codes-and-principles-ai
- De Jong, R., Koolm L. and Van Est, R. (2019) This is how we put AI into practice based on European Values, Rathenau Institute, The Hague, available at: https://www.rathenau.nl/en/digital-society/how-we-put-ai-practice-based-european-values
- Reis, W. C., Bonetti, A. F., Bottacin, W. E., Reis, A. S., Jr, Souza, T. T., Pontarolo, R., Correr, C. J., & Fernandez-Llimos, F. (2017). Impact on process results of clinical decision support systems (CDSS) applied to medication use: overview of systematic reviews. Pharmacy practice, 15(4), 1036. https://doi.org/10.18549/PharmPract.2017.04.1036

Medicine / Health information technology / Health Impact Assessment

- Montani, S. and Striani (2019), Artificial Intelligence in Clinical Decision Support: a Focused Literature Survey in: Yearbook of Medical Informatics.
- Sutton, Reed T. et al. (2020), An overview of clinical decision support systems: benefits, risks, and strategies for success in: npj Digital Medicine 3.
- Wasylewicz, A. T. M. and Scheepers-Hoeks, A. M. J. W. (2018), Fundamentals of Clinical Data Science (chapter 11 Clinical Decision Support Systems).
- Verughese, J et al. (2017), Cost and Economic Benefit of Clinical Decision Support Systems (CDSS) for Cardiovascular Disease Prevention: A Community Guide Systematic Review in: Journal of the American Medical Informatics Association 24 (3), 669–676, DOI: https://doi.org/10.1093/jamia/ocw160
- Shahsavarani, A. M. et al. (2015) Clinical Decision Support Systems (CDSS): State of the art Review of Literature in: International Journal of Medical Reviews 2 (4), 299-308.
- Lysaght, T., Lim, H.Y., Xafis, V. et al. AI-Assisted Decision-making in Healthcare. ABR 11, 299–314 (2019). https://doi.org/10.1007/s41649-019-00096-0
- Gheeshan, H., Malkhawi, O. and Khalaf, I. (2009). Computerized Clinical Decision Support Systems and their Clinical Application in Health Care Delivery System in: Jordanian Medical Journal 43, 267-273.
- Coeckelbergh, M. (2013) E-care as craftsmanship: Virtuous work, skilled engagement, and information technology in health care. Medicine, Health Care and Philosophy 16(4): 807–816.
- King Jr, B. (2018) Artificial intelligence and radiology: what will the future hold? In: Journal of the American College of Radiology 15 (3 Part B), 501–503.
- Krittanawong, C. (2018) The rise of artificial intelligence and the uncertain future for physicians. European Journal of Internal Medicine, 48, 14. doi:10.1016/j.ejim.2017.06.017.
- Mitchell, C. and Ploem, C. (2018) Legal challenges for the implementation of advanced clinical digital decision support systems in Europe. Journal of clinical and translational research, 3 (Suppl 3), 424–430.
- Lehne, M., Sass, J., Essenwanger, A. et al. (2019) Why digital medicine depends on interoperability. npj Digit. Med. 2, 79. https://doi.org/10.1038/s41746-019-0158-1
- Zikos, D. and DeLellis, N. (2018) CDSS-RM: A clinical decision support system reference model. BMC Medical Research Methodology. 18. 10.1186/s12874-018-0587-6.
- Finlayson, S. G. et al (2019) Adversarial attacks on medical machine learning, Vol. 363, Issue 6433, 1287-1289. DOI: 10.1126/science.aaw4399

- Panch, T., Mattie, H. and Celi (2019) L.A. The "inconvenient truth" about AI in healthcare. npj Digit. Med. 2, 77. https://doi.org/10.1038/s41746-019-0155-4
- McCallie D.P. (2016) Clinical Decision Support: History and Basic Concepts. In: Weaver C., Ball M., Kim G., Kiel J. (eds) Healthcare Information Management Systems. Health Informatics. Springer, Cham.
- Wright, A et al. (2010). Best Practices in Clinical Decision Support: the Case of Preventive Care Reminders. Applied clinical informatics, 1(3), 331–345. https://doi.org/10.4338/ACI-2010-05-RA-0031
- Pasar, A. (2019) Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes, Springer Science: New York.
- Obermeyer, Z. (2019), Dissecting racial bias in an algorithm used to manage the health of populations, Science 25, Vol. 366, Issue 6464, 447-453. DOI: 10.1126/science.aax2342
- Ford, R. A. and Price II, W. N. (2016), Privacy and Accountability in Black-Box Medicine 23 Mich. Telecomm. & Tech. L. Rev. 1. Available at SSRN: https://ssrn.com/abstract=2758121
- Dagliati, A. et al. (2018) Big Data as a Driver for Clinical Decision Support Systems: A Learning Health Systems Perspective in: Frontiers in Digital Humanities, https://doi.org/10.3389/fdigh.2018.00008
- Bright, T. et al. (2012). Effect of Clinical Decision-Support Systems: A Systematic Review. Annals of internal medicine. 157. 10.1059/0003-4819-157-1-201207030-00450.
- Jia P et al. (2016) The Effects of Clinical Decision Support Systems on Medication Safety: An Overview. PloS ONE 11(12): e0167683. doi:10.1371/journal.pone.0167683
- Moja, L. et al. (2014). Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. American journal of public health, 104(12), e12–e22. https://doi.org/10.2105/AJPH.2014.302164.
- Murphy E. V. (2014). Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. The Yale journal of biology and medicine, 87(2), 187–197.
- J. Varghese et al. (2018). "Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review". Journal of the American Medical Informatics Association. 25 (5): 593–602.
- A.D. Black. Et al. (2011). "The impact of ehealth on the quality and safety of health care: A systematic overview". PLOS Medicine. 8 (1).

Law

- Price II, W. N. (2015), Black Box Medicine in: Harvard Journal of Law & Technology Volume 28, Number 2.
- Hanson, J. (2018) Precautionary Principle: Current Understandings in Law and Society, in: Encyclopedia of the Anthropocene, Eds: Dominick A. Dellasala, Michael I. Goldstein, Elsevier, 361-366, https://doi.org/10.1016/B978-0-12-809665-9.10451-3.
- Mazur, J. (2019) Automated Decision-Making and the Precautionary Principle in EU Law. Baltic Journal of European Studies. 9. 3-18. 10.1515/bjes-2019-0035.
- Smuha, N. A. (2019), From a 'Race to AI' to a 'Race to AI Regulation' - Regulatory Competition for Artificial Intelligence, available at SSRN: https://ssrn.com/abstract=3501410 or http://dx.doi.org/10.2139/ssrn.3501410

- Wachter, S.; Mittelstadt, B. and Floridi, L. (2017), 'Why a right to explanation of automated decision-making does not exist in the general data protection regulation,' International Data Privacy Law, vol. 7, no. 2, 76–99 https://doi.org/10.1093/idpl/ipx005.

Policy studies

- Castro, D. and McLaughlin, M. (2019), Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence, https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence

- Schomberg, R. (2013). A Vision of Responsible Research and Innovation. 10.1002/9781118551424.ch3.
- Hunter, A. P. et al (2018), Artificial Intelligence and National Security – The Importance of the AI Ecosystem, Centre for Strategic and International Studies.
- Renda, A. and Simonelli, F. (2019) Study supporting the interim evaluation of the innovation principle, Centre for European Policy Studies.
- European Commission for the efficiency of Justice (CEPEJ, 2018), European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment
- Council of Europe, Committee of Experts on Internet Intermediaries (MSI-NET) (2018), Study On The Human Rights Dimensions Of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications.
- High-Level Expert Group on AI (AI HLEG, 2019) Ethics Guidelines for Trustworthy AI.
- High-Level Expert Group on AI (AI HLEG, 2019) A definition of AI: Main capabilities and scientific disciplines.
- High-Level Expert Group on AI (AI HLEG, 2019) Policy and Investment Recommendations for Trustworthy AI.
- European Parliamentary Research Service (EPRS/STOA, 2019), Artificial Intelligence ante portas: Legal & ethical reflections, Briefing European Parliament.
- European Parliamentary Research Service Scientific Foresight Unit (STOA, 2016), Ethical Aspects of Cyber-Physical Systems: Scientific Foresight study.
- Expert Group on Liability and New Technologies New Technologies Formation (2019), Liability for Artificial Intelligence and other Emerging Digital Technologies.
- M. Weda et al. (2018) Digitale beslissingsondersteuning in de zorg: Een verkenning, RIVM.
- Engler, A. (2020) A guide to healthy skepticism of artificial intelligence and coronavirus, The Brookings Institution. https://www.brookings.edu/research/a-guide-to-healthy-skepticism-of-artificial-intelligence-and-coronavirus/

Risk governance / risk assessment

- EPFL IRGC (2018). The Governance of Decision-Making Algorithms. Lausanne: EPFL International Risk Governance Center.
- IRGC (2018) Guidelines for the Governance of Systemic Risks, Lausanne: International Risk Governance Center (IRGC).

Other

- Perez, C. C. (2019), Invisible Women: Exposing Data Bias in a World Designed for Men.
- Dorst, H., J. Deuten and E. Horlings (2016), The Dutch science system in the European Research Area, The Hague, Rathenau Institute. https://www.rathenau.nl/en/vitale-kennisecosystemen/dutch-science-system-european-research-area

Non-scientific literature

- Steger, A. (2019) What Happens to Stolen Healthcare Data? Health Tech Magazine, Available at: https://healthtechmagazine.net/article/2019/10/what-happens-stolen-healthcare-data-perfcon
- Kobie, N. (2019) Everyone should be worried by big tech's huge NHS data grab, WIRED available at: https://www.wired.co.uk/article/google-apple-amazon-nhs-health-data.
- Bresnick, B. (2018) Top 12 Ways Artificial Intelligence Will Impact Healthcare, Health It Analytics, available at: https://healthitanalytics.com/news/top-12-ways-artificial-intelligence-will-impact-healthcare.
- Strickland E. (2019), AI Agents Startle Researchers With Unexpected Hide-and-Seek Strategies, Institute of Electrical and Electronics Engineers, available at: https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/ai-agents-startle-researchers-with-unexpected-strategies-in-hideandseek
- Hao, K. (2019) This is how AI bias really happens—and why it’s so hard to fix, MIT Technology Review, available at: https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/
- Etzioni, O. (2018) A Hippocratic Oath for artificial intelligence practitioners, TechCrunch, available at: https://techcrunch.com/2018/03/14/a-hippocratic-oath-for-artificial-intelligence-practitioners/
- Molyneux, C. G. and Oyarzabal, R. (2017) What is a Robot under EU Law? Global Policy Watch Key Public Policy Developments Around the World, Covington and Burling LLP, available at: https://www.globalpolicywatch.com/2017/08/what-is-a-robot-under-eu-law/
- Beatie, A. (2019), ‘Technology: how the US, EU and China compete to set industry standards’, Financial Times, available at: https://www.ft.com/content/0c91b884-92bb-11e9-aea1-2b1d33ac3271
- Galeon, D. (2017) Stephen Hawking: “I Fear That AI May Replace Humans Altogether”. Futurism. Available at: https://futurism.com/stephen-hawking-ai-replace-humans
- Helbing, D. et al. (2017) Will Democracy Survive Big Data and Artificial Intelligence? In: Scientific American, https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/
- McKinsey Global Institute (2018), Notes from the AI Frontier: Modelling the Impact of AI on the World Economy.
- Future of Life Institute (visited 9 April 2020), AI Policy China, https://futureoflife.org/ai-policy-china/?cn-reloaded=1
- Future of Life Institute (visited 9 April 2020), AI Policy United States, https://futureoflife.org/ai-policy-china/?cn-reloaded=1

- Future of Life Institute (visited 9 April 2020), An Open Letter RESEARCH PRIORITIES FOR ROBUST AND BENEFICIAL ARTIFICIAL INTELLIGENCE, https://futureoflife.org/ai-open-letter
- OpenClinical, Decision Support Systems, http://www.openclinical.org/dss.html (last modified 2013).
- Creighton, J. (2018) The "Father of Artificial Intelligence" Says Singularity Is 30 Years Away, Futurism.

Guidelines from organizations and companies

- Del Castillo, A. P. (2019) Foresight Brief: Labour in the age of AI: why regulation is needed to protect workers, European Trade Union Institute (ETUI).
- https://www.philips.com/a-w/about/news/archive/blogs/innovation-matters/2020/20200121-five-guiding-principles-for-responsible-use-of-ai-in-healthcare-and-healthy-living.html
- https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf
- Philips (2020), Five guiding principles for responsible use of AI in healthcare and healthy living. Available at: https://www.philips.com/a-w/about/news/archive/blogs/innovation-matters/2020/20200121-five-guiding-principles-for-responsible-use-of-ai-in-healthcare-and-healthy-living.html

EU legislation and regulation

- European Parliament, Digital Agenda for Europe, Fact Sheets on the European Union, https://www.europarl.europa.eu/factsheets/en/sheet/64/digital-agenda-for-europe
- https://ec.europa.eu/eurostat/cache/infographs/ict/bloc-4.html
- https://ec.europa.eu/digital-single-market/en/artificial-intelligence
- Eur-Lex, Human rights, https://eur -lex.europa.eu/summary/chapter/human_rights.html?root_default=SUM_1_CODED%3D13
- https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation
- https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html
- European Group on Ethics in Science and New Technologies (2018) Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems
- ICDprecautionary principleC (2018), Declaration on Ethics and Data Protection in Artificial Intelligence.
- EU, Declaration: Cooperation on Artificial Intelligence.
- COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS
- https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/2
- https://www.europarl.europa.eu/news/en/press-room/20200206IPR72015/artificial-intelligence-meps-want-to-ensure-a-fair-and-safe-use-for-consumers
- https://ec.europa.eu/info/research-and-innovation/law-and-regulations/innovation-friendly-legislation_en

Cases

- http://curia.europa.eu/juris/document/document.jsf;jsessionid=9ea7d0f130d5e6c8aa7ba9904c72b8bbb7753076a873.e34KaxiLc3eQc40LaxqMbN4PaNmLe0?text=&docid=197527&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=825346
- https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:86

Legislation

- Consolidated version of the Treaty on the Functioning of the European Union - PART THREE: UNION POLICIES AND INTERNAL ACTIONS - TITLE XIV: PUBLIC HEALTH - Article 168 (ex Article 152 TEC), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12008E168.
- Low Voltage Directive
- Electromagnetic Compatibility Directive
- The Radio Equipment Directives.
- DIRECTIVE 2011/24/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 9 March 2011 on the application of patients' rights in cross-border healthcare
- General Data Protection Regulation (GDPR), https://gdpr-info.eu/.
- European Commission (2020), White Paper on Artificial Intelligence - A European approach to excellence and trust.

9 Appendix

N/A