



HAL
open science

Too broad to handle: can we "fix" harmonised standards on artificial intelligence by focusing on vertical sectors?

Mélanie Gornet

► To cite this version:

Mélanie Gornet. Too broad to handle: can we "fix" harmonised standards on artificial intelligence by focusing on vertical sectors?. 2024. <hal-04785208>

HAL Id: hal-04785208

<https://hal.science/hal-04785208v1>

Preprint submitted on 15 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Too broad to handle: can we “fix” harmonised standards on artificial intelligence by focusing on vertical sectors?

Mélanie Gornet
Télécom Paris – Institut Polytechnique de Paris
melanie.gornet@telecom-paris.fr

Abstract

The European approach to regulating Artificial Intelligence (AI) has relied on three main regulatory mechanisms: ethics charters, the AI Act and technical standards. Europe has based this approach on concepts such as “trustworthiness” or “risk”, navigating a semantic sphere where the ethical, legal and technical fields clash. The origins of this approach in ethics charters, which usually focus on broad principles, have led to the dissemination in the AI Act and in standards of a very general discourse about AI, which rarely goes into technical detail, and with elements that are unimplementable as is. Additionally to this broadness of principles and requirements, the European discourse on AI, whether in ethics charters, the AI Act or standards, has also remained very horizontal. While the AI Act classifies high-risk systems according to their sector of use, the obligations applicable to them are the same regardless. This poses a problem for standards, which are forced to remain at a high level, as the technical requirements are too difficult to define without contextual elements. We therefore propose to refocus standards on vertical sectors, allowing them to define stricter requirements.

1 Introduction

AI regulation has been a topic of interest for the last fifteen years, with the publication of various ethics charters, legal texts and technical requirements around the world. This movement reached Europe in the late 2010s, with the publication of the High Level Expert Group (HLEG) guidelines (HLEG, 2019) and, more recently, of the AI Act (European Parliament and Council, 2024a), the first mandatory framework for AI in the world. Since then, Europe has refined its approach based on “trust”, a philosophical concept that has become a marketing tool. In the AI Act, providers of high-risk AI systems will have to respect a number of obligations, often by testing the technical properties of their systems (art. 9 to 15). But the details of these requirements are not defined in the AI Act but in technical standards. Subsequent efforts have therefore focused on developing a certain type of standards, harmonised standards (hENs), which are technical documents drafted by European Standardisation Organisations (ESOs), and in particular by CEN-CENELEC¹ JTC² 21 in the case of AI systems, at the request of the European Commission. These hENs, if they are cited in the Official Journal of the European Union (OJEU), will be granted legal effects (Gornet and Maxwell, 2024). This includes the presumption of conformity with the law, when a provider is compliant with the respective hENs. This approach to product safety is common in Europe since the 80s, and is called the New Legislative Framework (NLF).

¹European Committee for Standardisation (From French: Comité Européen de Normalisation) and European Committee for Electrotechnical Standardisation (From French: Comité Européen de Normalisation Électrotechnique).

²Joint Technical Committee.

This three-step approach to AI regulation in Europe has, however, generated much discussion in the literature because of its distinctive features. Indeed, the AI Act mixes rights and risks and therefore asks standards to address fundamental rights issues for which they are not equipped (Gornet and Maxwell, 2024). Further discussions with standardisation experts also revealed that the horizontal and cross-cutting approach of AI standards, inherited from ethical charters and the AI Act, is difficult to implement in practice (Gornet and Herman, 2024). What are the problems raised by the European approach to AI regulation and can we suggest improvements?

This piece is a position paper, where we rely both on the analysis of previous literature and European institutions documents, as well as on a previous study (Gornet and Herman, 2024), where we interviewed 16 experts from CEN-CENELEC JTC 21. Their names have been anonymised and have been replaced by P1 to P16. When their words have been translated from a language other than English, the quote is marked by an asterisk (*). For more information about the interview protocol and the experts' background, see (Gornet and Herman, 2024).

We begin in Section 2 by examining ethics charters, the AI Act and technical standards, how they each draw inspiration from the other two, and we show that this is not necessarily to their advantage. In Section 3, we look at the content of standards and show that, although they were originally intended to define specific technical requirements, with AI, they tend to focus on governance and normative issues and we question whether this is the right way forward. In Section 4, we recall the problems that the standardisation system is facing, from the restricted access to standards, the influence of international voices and the private sector, to their general lack of legitimacy. We show that this is mainly due to a shift in standardisation, towards the inclusion of more diverse stakeholders at the request of the Commission, the horizontality of standards and the strict deadlines to which they are subject. We therefore propose in Section 5 an alternative solution to the current AI standardisation strategy: refocusing standards on vertical sectors, where they could define more specific technical requirements and appeal to a more diverse audience. However, this will not come without some cost, given that there is no more time before the end of the deadline for harmonised standards, and standards for the AI Act are expected to be horizontal.

2 Ethics, law and standards constantly overlap

2.1 Technical and legal influences on ethics: a loss of meaning

Today, with AI in particular, ethics is wrongly applied to a technological object. In its traditional sense, ethics is associated with moral behaviour (Merriam-Webster Dictionary, n.d.) and sets of principles intended to guide a person's actions. It therefore applies to reasoning or a moral being and, like trust, cannot become a characteristic of AI.

Nevertheless, AI ethics charters have become accustomed to defining these principles in the form of checklists, which resemble more good development practices than virtuous behaviour. This technicisation of ethics is observed by (Hagendorff, 2020), who explains that technical elements are conceptualised in AI ethics charters as *“isolated entities that can be optimised by experts so as to find technical solutions for technical problems”*. There is therefore a risk of diversion from the traditional conception of ethics³.

In addition, this approach to ethics as a set of technical properties is also characteristic of what Hunyadi (2015) calls “Small Ethics”⁴. “Small Ethics” is defined by the author as *“the liberal ethic of individual rights and freedoms”* (Hunyadi, 2018), it is small not because it is not important

³What (Tessier, 2022) calls, in French, *“le dévoiement de l'éthique”*.

⁴From French, *“Petite éthique”*, although the translation “Small Ethics” is used by the author himself (Hunyadi, 2018).

but because it is centered around the individual rather than having a global perspective. In the context of new technologies, “Small Ethics” is part of a “fait accompli” logic (Tessier, 2022), where individuals must adhere to these ethics principles and the lifestyle they convey, thereby restricting their freedoms. This is also called “*ethics of carefulness*”, or “*checklist ethics*” (Bolte et al., 2022).

As part of the EU’s approach to AI regulation, ethics guidelines, such as those from the HLEG, were also a preparatory step to mandatory legislation, preparing concepts such as “trustworthiness” and “risk”, which would be reinforced in the AI Act. However, ethics frameworks are meant to complement legislation, not prepare it. Ethics should not be confused with a preparatory stage for mandatory legislation, nor should it be constrained by law. On the contrary, it should try to go further than the law, because respect for the law is the minimum requirement in a democratic society.

AI ethics charters are therefore used for different purposes, as technical checklists or as a basis for regulation under the European approach. However, these influences are detrimental to the essence of ethics, which is twisted in the current frameworks and loses some of its meaning.

2.2 Ethical and technical influences on law: the European strategy based on keywords

AI projects and initiatives, whether research articles, policy documents or industry communications, tend to use catchy keywords to surf on the hype of AI ethics (Probasco et al., 2023). Different spheres of AI stakeholders are publishing their own charters or policy documents, using a rather convergent discourse, but which still conceals significant divergences, specifically on the vocabulary that is used. A potential explanation to this mix of discourses is the influence of both the industry and the civil society on governmental and supra governmental entities which become arenas to push one’s interests.

Since the High-Level Expert Group (HLEG) guidelines in 2019, the European strategy for regulating AI has been based on a notion of “AI trustworthiness”. Indeed, for the HLEG experts, trust is considered to be the “*bed-rock of societies*” (HLEG, 2019, p.4). For Laux et al. (2023), citizens are encouraged to trust AI so that they can use it more and businesses and governments can reap its economic benefits. Indeed, studies have shown that greater trust often translates into positive economic development (Bjørnskov, 2017). There is therefore a deliberate confusion, in the AI Act, between “trustworthiness” and “acceptability” (Laux et al., 2023).

The HLEG defines “trustworthiness” as a concept based on three pillars: ethical AI, lawful AI and robust AI. But this definition is far from being a given in the literature on trust. For Nickel et al. (2010), there are two types of trust: the pure “*rational-choice*” notion of trust, which cannot be differentiated from the engineering notion of “reliability”, and the “*motivation-attributing*” notion of trust, which require the trustor to attribute some motivations to the trustee with regards to their values and interests. In that last sense, AI cannot be “trustworthy” as it is not a human agent, and has no moral or motivations. Ryan (2020) adds that AI does not have the capacity to be trusted because it cannot held responsibility for its actions. For Ryan (2020), speaking of “trustworthy AI” is therefore a dangerous anthropomorphism, which wrongly attributes moral and emotional capacity to technology. But some authors still think an adaptation of the notion of trust is possible to fit the AI context. Rieder et al. (2021) proposes that an AI system could be considered “trustworthy” when it is responsive to the interests of the human trustors. As such, it does not need to have its own motivations, but simply to know those of the human party. Yet, despite a few exceptions, most of the literature agrees that AI cannot be “trustable”, although it can be “reliable” (Laux et al., 2023).

In the context of the AI Act and AI standards, standardisation experts give a definition of

“trustworthiness” that is indeed closer to that of “reliability”. One expert from CEN-CENELEC JTC 21 explains: “We are working on ‘trustworthiness’ according to the European Community’s approach, which is a market approach. The market approach to ‘trustworthiness’ allows for possibilities based on certain characteristics”*[P2]⁵. These characteristics are the technical requirements that will be defined in standards to precise the essential requirements of the AI Act. This definition is coherent with the one given by international standards on AI where trustworthiness is defined as “the ability to meet stakeholders expectations in a verifiable way”⁶ (ISO/IEC, 2020). On the contrary, reliability is defined as the “property of consistent intended behaviour and results” (ISO/IEC, 2020). While the two are clearly different according to standards, there are still inconsistencies that are revealed by translation into different languages. For instance, the “overview of trustworthiness in AI” proposed in ISO/IEC 24028 (2020) is translated in French by something much closer to “overview of reliability”⁷.

We are therefore witnessing a semantic shift whereby a discourse on “ethics” has been transformed into a discourse on “trustworthiness”, which, in the HLEG guidelines, is supposed to encompass ethics, law and technical properties, but which is gradually being understood and reduced to the latter dimension in the AI Act and in standards. This shift is justified by the German Institute for Standardisation (DIN) because, in their view, ethics refers to rational beings and trustworthiness could refer to organisations and technical systems (DIN and DKE, 2022). However, the emphasis in Europe on trustworthiness rather than reliability appears to be part of a wider communication strategy to make AI more acceptable. The same conclusions can be drawn from the focus on “risk” where, apart from a few unacceptable ones, many risks are considered acceptable under the right conditions.

As such, the EU’s approach to AI regulation with, in particular, the AI Act, which is based on a specific version of trustworthiness, advertise its ethical origins, but is actually more influenced by technical notions of reliability.

2.3 Legal and ethical influences on standards: are standards more than industry documents?

Technical standards for AI are increasingly trying to address normative questions, which are usually reserved for ethics charters (Gornet and Maxwell, 2024). This development is primarily the result of the European Commission’s desire to address the risk to fundamental rights in its AI Act. As a result, the ESOs have adopted a horizontal and high-level approach to AI standardisation, taking up the semantics of the AI Act in an upcoming standard called the “AI trustworthiness framework”. According to a dashboard released by CEN-CENELEC experts on social media mapping JTC 21 standards to the requirements of the AI Act (CEN-CENELEC, 2024), the “AI trustworthiness framework” is intended to cover seven of the ten items present in the standardisation request (European Commission, 2023), which makes it very broad and horizontal, a characteristic quite uncommon for technical standards (Gornet and Herman, 2024). This trustworthiness framework which, like previous ethics charters, attempts to address technical criteria of reliability.

At the same time, standards have taken on a legislative role in Europe with the New Approach since the 80s. The combination of more normative issues and increased legal powers brings these frameworks closer to mandatory legislation. However, standards are advertised as documents

⁵This quote, as well as other quotes in this work come from the interview we conducted in a previous study (Gornet and Herman, 2024).

⁶Stakeholder expectations include accountability, accuracy, availability, controllability, integrity, quality, reliability, resilience, robustness, safety, security, transparency, and usability.

⁷“Examen d’ensemble de la fiabilité en matière d’intelligence artificielle”, instead of “Examen d’ensemble de la confiance”.

drafted by so-called “experts”, i.e. individuals who have technical and industrial expertise and who are capable of implementing these specifications in a business pipeline. There is therefore a dissonance between what standards fundamentally are and what they are becoming in the context of European regulation. This dissonance notably leads to problems of legitimacy that the European institutions are trying to resolve. But standardisation bodies are struggling with this new role that is not tailored to their organisational structure.

2.4 Interacting without blending

Ethics has lost some of its meaning in the attempt to define more technical criteria in charters; the AI Act bases its entire strategy on “trustworthiness”, inspired by ethical-technical frameworks; and standards, which were already considered legal acts under EU law, are becoming more normative with AI. As a result, the three spheres that the European AI regulation strategy is trying to address are actually quite intertwined, with increasingly blurred borders.

But with this blending, each of these frameworks actually loses its effectiveness and legitimacy, because the people who draw them up cannot be competent in all areas. According to [Eliantonio and Medzmariashvili \(2017\)](#): “*EU law, while embracing hybridity, has not yet found a way to come to terms with it*”. There are therefore calls to keep these spheres more separate. The European Parliament has notably stated, several years ago that “*it is of the utmost importance to draw a clear line between legislation and standardisation in order to avoid any misinterpretation with regard to the objectives of the law and the desired level of protection*” ([European Parliament, 2010](#)). Similarly, the mix of technical standards and fundamental rights satisfies neither the standardisation experts nor civil society associations⁸.

Regulatory frameworks that try to do too much run the risk of doing nothing properly. We therefore recommend avoiding this mix of approaches as much as possible and focusing each framework on what it does best: ethical charters to guide reflection and action, the law to protect rights and standards for specific technical elements. These frameworks should still interact, but all the while having their own separate scope.

3 Technical standards are not that technical

Standards are becoming less and less technical. This is due to two trends: (i) making standards more about management processes than about the products themselves, what [Gornet and Maxwell \(2024\)](#) have called “governance” standards, and (ii) incorporating ethical, social and even fundamental rights issues into standards. But what can we learn from these trends and are they the right approach to standardising AI?

3.1 Addressing governance versus technical properties

Standards for AI fell into two main categories: either “governance” or “information” standards ([Gornet and Maxwell, 2024](#)). Governance standards concern the processes put in place by companies to produce AI systems, while information standards define technical terms and properties associated with AI systems. However, contrary to other fields, there is no “performance” standards for AI at the moment, i.e. standards setting minimum thresholds for certain properties of AI systems. But which of these types of standard is best for standardising AI?

⁸See Section 3.2.

Governance standards are flexible

Governance standards include quality management and risk management standards. They are more concerned with what companies can do and what mechanisms they can put in place to prevent something from going wrong than with checking whether there is in fact an issue with the product. The purpose of governance standard is very different from that of product standards⁹, so they cannot easily be compared. It is the most widespread type of standard in the modern era (Yates and Murphy, 2019), particularly for new technologies where agreement on the constantly evolving state of the art is virtually impossible (Choi, 2024). This trend is, however, fairly recent. It developed in the 1980s with quality management standards and software standards and is now becoming the main type of standard for AI.

Governance standards are notably criticised for not being strict enough and for leaving companies too much room to manoeuvre. Indeed, companies often do not need to change their existing processes, and compliance to the standard is used as a justification for bad practices. As [P1] puts it: “[auditors] simply check that the company is achieving the objectives it has set itself”*. This lack of control can be dangerous if it is set up as a mark of quality used as a marketing tool by companies, but it is even more dangerous when it becomes the only safeguard required by law. These standards have also been criticised for not being adapted to SMEs or startups. For example, the ISO/IEC 42001 standard, on risk management for AI (ISO/IEC, 2023), deals with the organisation of the company and requires the presence of several teams – legal, technical, etc – a structure that a small startup simply cannot afford. Rendering the standard almost compulsory by harmonising it could therefore kill off small businesses. In that regard, the European Commission, which works closely with standardisation organisations, including CEN-CENELEC, has already stated that ISO/IEC 42001 was not enough for supporting the AI Act risk-based requirements. Indeed, the AI Act requires providers of high-risk AI systems to have a risk management framework in place, a framework that will be standardised by future hEN. ISO/IEC 42001 has therefore been adopted by CEN-CENELEC and has become a European standard, but is not in the process to become a harmonised standard to support the AI Act.

But just because ISO/IEC 42001 is not sufficient to ensure a sufficient level of protection for Europe does not mean that governance standards are useless. Particularly in the context of the AI Act, it will be necessary to put in place a risk management framework in companies providing high-risk systems. These frameworks will need to be standardised as the AI Act relies heavily on standards to implement them. Without effective risk management standards, the legal text would risk being unenforceable. Yet Europe needs to be cautious about the level of requirements that these standards convey, particularly if they are to support the AI Act. In particular, they must be applicable to all companies and be sufficiently strict not to justify bad practices and enable effective supervision. Above all, they must not be used as the sole means of protection but they can complement more technical standards, such as information standards.

Information standards avoid technical details

Information standards are often more technical than governance standard. Even if they do not set thresholds, the technical definitions and benchmarking possibilities they offer can push companies to innovate and compete to achieve a higher level of protection. This protection includes security measures as well as measures relating to fundamental rights¹⁰. However, even when it comes to definitions and measurements, the question arises as to how far standards should go into technical

⁹Both information and performance standards.

¹⁰See Section 3.2.

detail. Indeed, the standardisation of techniques that are not sufficiently mature has proved to be a considerable brake on innovation in several areas prior to AI. This is why some experts claim during standard development that a measure cannot be included in a standard, because although it is widely used in the field, its effectiveness has not been proven [P5]. But at the same time, not going into the technical details of the technology makes a standard virtually useless. This is why some standards experts are in favour of including more definitions and measurements in standards in general, at the risk of them becoming obsolete in a few years' time. Others, are more cautious about what they want to include, at the risk of having standards that are emptier and further removed from the current state of the art. The level of maturity of technical properties, tests or measurements that is sufficient to incorporate them into a standard is therefore relative and not an exact science.

At the same time, it is interesting to examine the dynamics of the pressures exerted by the various parties in the context of standardisation. BigTech companies are often in favour of including less technical elements in standards, either by developing more governance standards or by making definitions and measurements in information standards less technical.

There are many examples of how companies can undermine mandatory regulation by refusing to include technical elements, such as precise metrics and mathematical formula, in standards, preferring broad and sometimes circular definitions. Examples include a definition of transparency that does not refer to the notion of access to the system for audit purposes, or a definition of data representativeness that is unimplementable (Gornet and Herman, 2024). This imprecision in standards, even the most technical ones, is due to the fact that standards, as a form of regulation, limit companies in their practices while companies would prefer to be as free as possible. Market dynamics also have a major influence on these positions, particularly for companies that intend to bring to market new innovative products that are technically far removed from what is described in the standards.

3.2 Addressing ethical, social and fundamental rights questions

A push for more normative questions in standardisation

The question of the appropriate technical level of AI standards should be supplemented by another question: should ethical, social and fundamental rights issues be standardised? And if so, how? Indeed, the European Commission, with the AI Act, pushes the standardisation ecosystem to integrate these issues into standards. Even before the AI Act, ethics was cited as a potential subject of interest for standardisation organisations (CEN-CENELEC, 2020; DIN and DKE, 2022). But with the AI Act, Europe has been propagating a dual discourse, which aims to protect fundamental rights while at the same time having a risk-based structure inspired by product safety regulations. This has led the ESOs to develop a variety of standards which deal to some extent with ethical and fundamental rights issues, such as the standard on AI “trustworthiness”, and other initiatives, on the competencies of AI ethicists or how to conduct a fundamental rights impact assessment for instance. But Europe is not alone in its journey to standardise ethical, social and fundamental rights questions. Other initiatives have been launched by ISO, IEEE¹¹ and other organisations (Gornet and Maxwell, 2023).

¹¹Institute of Electrical and Electronics Engineers.

Not everyone agrees that these questions should be standardised

However, addressing ethical, social and fundamental rights questions in standards can be hard. Apart from the Commission, other European institutions are cautious about incorporating these issues into standardisation, with the European Parliament stating that standards should not address “*fundamental rights or socio-economic issues*” (European Parliament, 2023). Some oppose these initiatives, even within standardisation bodies, claiming that ethics is not “tangible” enough to be standardised or that it is not the role of the ESOs to standardise fundamental rights. This type of statement seems to be common in engineering task forces and standardisation groups, with experts claiming that they “*do not do politics*”, or that they are “*just engineer[s]*” (Cath-Speth, 2021). Similarly, among the experts we interviewed, those with a technical background admit that they are not really interested in ethics [P15]. In a recent study (Baeva et al., 2023), a standardisation expert working on AI explained: “*We cannot write into a standard what is good and what is bad. What we can write into it, however, is a description of the ethically relevant properties of the system, i.e., the degree of transparency, the degree of fairness, the degree of privacy, the degree of robustness, and the methods for measuring them.*” This statement seems to be in line with the EU’s approach to “trustworthiness”, which is closer to a list of technical properties that guarantee “reliability” than to ethical considerations.

There does not even seem to be a consensus on what constitutes such ethical and fundamental rights standards. Some standards, such as CEN-CENELEC “competency framework for AI ethicists”¹², are clearly considered by all to be “ethical standards”, but more technical standards, such as fairness standards, are not always present in everyone’s mind. The “AI Trustworthiness framework”, for instance, is not always considered by CEN-CENELEC experts as an “ethical” standard. We found during our interviews that experts seem to differentiate between “technical” standards and all the other standards, which deal with ethical, social or fundamental rights issues, but at a higher level. This last category is sometimes considered the least important, because the main objective of SDOs is generally to produce technical specifications, and the experts themselves sometimes feel less entitled to discuss these issues. Additionally to standardisation experts themselves, various civil society organisations that focus on defending citizens’ rights have similarly called for standards to be limited to technical aspects (EDRi, 2022; Giovannini, 2021).

Standards are not neutral

However, we do not subscribe to this categorisation of standards between “technical” and “everything else”. In our opinion, the boundary between what is ethical and what is technical is difficult to draw, resulting in standards that may make value judgements without the developers of such standards even realising it (Gornet and Maxwell, 2024). Fairness standards are a good example: deciding on a threshold for a specific measure of (un)fairness can be understood as setting a threshold for an acceptable level of discrimination, given that residual biases will always be present. With regard to performance standards, the choice of a threshold is a normative judgement, even for safety standards, putting a quantitative measure on the right to safety and sometimes even on the right to life of individuals. Safety thresholds, however, are more accepted and more widespread in our society than thresholds for fundamental rights. What makes fundamental rights different?

One of the main differences between safety standards and standards relating to ethics and fundamental rights is the level of subjectivity and cultural significance. For example, technologies that are accepted in some countries will not be accepted in others. In that regard, the EU has taken

¹²For a complete list of JTC 21 published standards and standards under development, see (CEN-CENELEC, n.d.b,n).

a firm stance by banning social scoring systems, which are already widely used in China, which accounts for 17% of the world’s population. There is therefore no consensus among the world’s population on what use of technology is “ethical” and respects fundamental rights. As with the choice of whether or not to use a technology, the choice of whether a risk to fundamental rights is acceptable or not is a normative and political choice that is codified in standards: what do we accept as a society – or as a coalition of countries, in the case of Europe?

But what differentiates the AI Act approach from that of standards is that the AI Act is a legal text which, as such, necessarily codifies certain aspects of ethical behaviour and defines the rights of individuals. The AI Act can set normative thresholds on what is acceptable or not because the EU institutions have the legitimacy to make these choices. On the contrary, standards are drawn up by private bodies whose work is supposed to remain technical. Despite their power within the EU due to their close collaboration with the European institutions, ESOs are not supposed to have a say in policy-making.

Finally, there is also the question of the maturity of the technology. If we accept a residual risk in the case of nuclear activity, this is both because the alternative would be to stop the activity altogether, which would otherwise benefit humanity, and because the “residual risk” is very low, the chances of a nuclear accident being of the order of one in billions. In the case of AI, it is questionable whether the use of AI systems is sufficiently beneficial to society to take significant risks to fundamental rights by using them. Additionally, the threshold that standards would have to set would be much higher. Indeed, the “residual risk” to fundamental rights is partly higher because AI technology is still in its expansion phase, where new advances are made very rapidly, and because of the lack of causality that makes AI systems behaviour difficult to predict.

The danger of setting thresholds

In our analysis, we have not seen any AI standards aimed at setting thresholds for the acceptability of risks to fundamental rights. This is not to say that it is not a possibility. In the case of fairness, as metrics develop, we could imagine one day having a standard that would draw the line between what is an “acceptable” level of biases and what is not. This threshold could be decided following different rationales: the standard could either fix an “olympics” threshold that do not settle for less than the best possible solution in the state of the art, or a “filter” threshold that only fixes an arbitrary limit (Busch, 2011). Each of these types of threshold allows a different version of what is “acceptable”.

In the case of a filter threshold, on the one hand, it would push industrial players to improve their technology, at least up to a certain point. On the other hand, it risks setting a threshold which, once reached, will slow down companies in their quest for improvement. Additionally, setting a filter threshold today would run the risk of setting limits that would be obsolete in a few years’ time. Even if standards can be updated, this constant lag behind technological advances risks causing a number of breaches to people’s fundamental rights, whereas the state of the art would make it possible to avoid many incidents.

Yet, when thresholds are set for criteria that may affect fundamental rights, even when they are olympics thresholds, they still allow certain harms to occur and to be permitted. For example, an AI system may still be discriminatory even when the provider has applied the most modern techniques to avoid bias. This discrimination could have happened either because the state of the art will never allow all bias to be eliminated, or because the system is used in a way that creates bias and discrimination¹³. Unlike a case where AI regulation would be rights-based, under the current AI Act, system providers who would have applied this olympics threshold would be excused because

¹³For example, the feedback loops that are created when the output of an AI system is used as input.

they have done everything they could to avoid discrimination. However, such a threshold would still be better than a filter threshold where companies would only have to tick the boxes set up by standards without thinking about the possible other harms their systems could cause.

Remaining cautious while standardising information

But even when standards do not try to set thresholds, addressing ethics and fundamental rights still remains difficult. In information standards, the definitions given and the metrics listed can have an influence on the tests carried out and therefore on the problems that are revealed, but above all on those that remain hidden. We have seen that the chosen measure of fairness, for example, can lead developers to conclude that the system is or is not biased. The conclusion on whether discriminatory practices have indeed taken place is therefore closely linked to this technical choice. Presenting developers with a list of metrics without forcing them to test them all would only allow them to choose the one that proves that their system is not discriminatory.

On the other hand, standards must necessarily define the technical state of the art, and not acknowledging the existence of these metrics in standards would simply sweep the problems of unfairness under the rug. Particularly for standards which will support the AI Act, the requirements that appear in these standards are likely to be the only ones implemented by companies, at least initially. As such, they need to be as technically precise and as complete as possible, without taking any decision for the developers. The final decision of which metrics to use and what test to carry should mostly be left out to the responsibility of the developers.

If AI standards confine themselves to defining either processes in governance standards, or technical definitions and metrics in information standards, the design choices made by the developer as to which test to perform will carry a normative value of what they consider to be a “sufficient” level of protection. It will then be up to the judges, in case of litigation, to decide whether these technical measures were appropriate or whether they infringed fundamental rights.

3.3 A question of balance?

The degree of precision of the metrics, tests and procedures that standards should include to ensure compliance with the requirements of the AI Act depends first and foremost on whether policymakers want a fully ex ante approach to AI regulation.

Indeed, if the aim is to prevent most damage before systems are put on the market, the right approach should probably be to define as many things as possible in standards, including olympics thresholds. This would also put an end to a great deal of damage, by making the systems safer and more respectful of fundamental rights. However, this would be to the detriment of the EU’s other objective, namely the free movement of goods and the desire not to hinder innovation. It could also be a heavy burden for smaller companies which will suffer from compliance costs. Additionally, residual problems that have not been defined in these standards will be rendered non-existent and it will be even more difficult to prove that anything has gone wrong. In short, this would make it possible to avoid certain prejudices to the detriment of others, which are less measurable and therefore less defined in standards, but which would nevertheless have a significant impact on fundamental rights. The biases of facial recognition systems are a case in point: until the outrage caused by academic studies and false accusations in the late 2010s, society was unaware that these systems could discriminate against groups of people. Although some of these biases have since been dealt with, there is no telling how many problems like this one persist in AI systems that we are unaware of. Giving too much power to standards that define the current state of the art could prevent the discovery of new problems. But above all, even if they are known, these problems will

not be addressed by companies because standards do not require them to be. Furthermore, there is no guarantee that, even if there is the political will to define very strict rules for companies, the ESOs, which are mostly run by these companies, will comply and develop very technical and restrictive standards.

On the contrary, standards now have to face up to the possibility of not defining too many technical requirements, firstly so as not to upset companies and jeopardise the European market, but above all to leave room for interpretability of what constitutes “sufficient” protection. Less stringent standards would have the disadvantage of leaving some damage unaddressed when systems are put on the market. But at the same time, leaving it up to companies to decide what they want to put in place to protect individuals makes these choices more open to scrutiny and ensures that developers take responsibility for their decision-making. This type of transparency and openness, where providers of AI systems would only have to adhere to the same harmonised practices of testing, without obligations on the results of these evaluations, is what [Laux et al. \(2024\)](#) call “*ethical disclosure by default*”. But is it enough?

Firstly, this approach, which is in line with the promotion of governance standards, is supported by some industry players as it would allow greater flexibility in the implementation of requirements, but it is also supported by some academics who believe that strong normative choices should not be made in standards but should be dealt with by the actors involved in conformity assessment ([Tartaro, 2024](#); [Gonzalez Torres and Ali-Vehmas, 2024](#)). What this solution neglects to say is that, under the AI Act, conformity assessment will be carried out mainly by the AI system providers themselves, who are sometimes the same people who wrote the standards.

Secondly, an approach based on simple transparency runs the risk of simply having suppliers carrying out audits in accordance with the harmonised practices set out in the standards, but who do not concern themselves with the potential harmfulness of the resulting system. This disclosure obligation is similar to what is required for AI used for recruitment in the state of New York in the US, with regard to algorithmic biases. However, studies have shown that although providers disclose the results of their bias audit, as required by law, these systems still contain many biases. Providers do not mitigate these biases because they are not obliged to do so, only to disclose them ([Groves et al., 2024](#)). Simply focusing the standards on disclosure and good management practices could have a similar effect if they do not require anything about the outcome of these assessments.

The question of what standards should contain, both in terms of their technical level and the way in which they address normative issues, is a political choice. The ecosystem of standardisation seem to tend to prefer governance standards that leave room to company to decide what process they want to implement. But as [Yates and Murphy \(2019\)](#) notes, the shift to quality management standards has led the world of standardisation to become “*ironically less focused on the social good*”. Indeed, large companies appreciate these open-ended standards because they often do not even need to change their processes to meet them, they simply have to generate documentation ([Choi, 2024](#)). If too much room is left to companies, without ensuring that the choices they make indeed guarantee a high level of protection to fundamental rights, the safeguards of standards and subsequently of the AI Act will fail.

3.4 Getting out of the neutrality discourse

The standardisation sector presents itself as an area of technical expertise, producing guidelines and requirements that are supposed to be based on science. Standardisation experts generally insist that technical considerations alone determine standards ([Loya and Boli, 1999](#)). Their legitimacy to define the best practices to follow is based on the fact that they are professionals who have

knowledge of the field and can be considered “technical experts”. But this reliance on expertise poses problems for three reasons. First, it fails to recognise that normative choices are embedded in science and technique. Second, we have seen that this expertise, at least in the case of AI, lies not so much in science itself, but rather in the processes of SDOs. Finally, these organisations prone governance principles that are erected as universal, but which itself does not properly respect.

In the words of [Loya and Boli \(1999\)](#), the authority of standards bodies “*derives from the truth presumably embodied in science and technique, from the righteousness presumably embodied in the principles of governance by which the standards bodies operate (equality, fairness, nonpartisanship), and from the presumed self-interest of the lower-level actors that comprise them.*” Standardisation is based on the same ideal of neutrality in science that scientists have long endorsed ([Harding, 1992](#)). This use of the neutrality argument is not new, as science has often be used in history to justify lobbying activity ([Saltelli et al., 2022](#)). Similarly, it is not unusual for standards setters to share the view that technology is non prescriptive. The discourse of the standardisation experts we interviewed bear witness to this, as there seems to be a mistrust of normative questions. Many experts fail to recognise that what they choose to include in standards is in itself value-laden. For [Winfield \(2019\)](#), there is no standard which is not an ethical standard. Failure to recognise that no technical standard is neutral runs the risk of experts not being sufficiently critical of the choices they make.

But this neutrality discourse is also present in the marketing of standards, where the expertise of the people who develop standards is often put forward. However, we have seen that many of the experts who develop AI standards are not computer science experts, but governance experts who are familiar with standardisation processes. In fact, some of them are working on standards in several fields at the same time – AI but also quantum computing, software, etc. There is a disconnect between what is displayed by SDOs as a selling point for standards and the reality of profiles in standards development. For standards outside of AI, one could argue that what matters is not technical expertise but field expertise. However, as far as AI is concerned, the discussions in standards setting are captured by cross-sector BigTech companies. This is coherent with the objective of AI standards today: to be horizontal and high level in order to be applicable to any AI technology. This pushes aside sector-specific considerations and places less value on field expertise, which depends on the context of deployment of AI technologies.

According to [Loya and Boli \(1999\)](#), the standardisation sector “*displays principles of universalism, rationality, and homogenisation to an extreme degree*”. This is best illustrated by the World Trade Organisation (WTO) principles for international standardisation processes ([WTO, 2000](#)), to which most SDOs adhere. These principles include for instance “transparency”, “openness” and “impartiality”, to which the same criticisms can be levelled as to the ethical principles present in AI ethics charters: their generality and imprecision make it very easy to adhere to them, but more difficult to agree on how to respect them in practice. SDOs have even been accused of failing to respect some of these principles, such as the principle of transparency, as the inner workings of these entities and their processes are still highly opaque. Vague principles and statements can also be found in the CEN-CENELEC code of conduct ([CEN-CENELEC, 2018](#)) that members and observers participating in the work of CEN-CENELEC must respect. These include “*work for the net benefit of the European community*”, “*behave ethically*” and “*respect others*”. Incorporating these principles into a code of conduct or international agreement enables the SDOs to protect themselves against criticism, for example by citing these documents when accused of opacity or partiality. Furthermore, the notion of consensus, which is enshrined in these two texts and is at the basis of the operation of many SDOs today, can also be used as an excuse when the content of the standards is criticised for being partial. Experts and organisations have argued that if an element is included in a standard, it is because it has been the subject of a consensus, which justifies its adoption and

legitimacy. However, what is not as visible is that this consensus is the work of a very small number of people, with a majority of big industrial stakeholders.

The neutrality discourse held by the SDOs and their experts can be harmful, as it conceals certain problems of standardisation, such as the normative choices made by experts, the influence of economic and political incentives on standards, the lack of sector-specific technical contributions and the overall dysfunctions of an opaque and partial system. Whether the content of standards is more technical or more focused on governance, it is important to recognise that the content of standards is biased in favour of the interests of experts, in order to understand how standardisation could be improved.

3.5 AI standards could mark the beginning of a fourth wave of standardisation

AI standardisation has elements from the third wave of standardisation

The history of the evolution of standardisation through the ages is made up of numerous waves and trends that have shaped the system we know today. In particular, what [Yates and Murphy \(2019\)](#) has called the “third wave” of standardisation, which dates from the 1980s and continues to this day, relies on four key elements¹⁴: (i) a diversification of standards setting organisations, with groups of professional organisations taking over the role of traditional organisations; (ii) a diversification of profiles within these organisations, with more representatives from the civil society; (iii) a diversification of types of standards, with more “quality management” standards and less “product requirements” standards; and (iv) a diversification of processes of standards setting, with a decision making process that resembles deliberative democracy.

At first glance, AI standardisation meets all the conditions to be considered as a type of third wave standardisation. Firstly, although traditional standards bodies such as ISO still have a strong presence in AI standardisation, particularly with the extensive work of SC 42, we are indeed seeing a number of other organisations setting standards for AI, particularly professional organisations such as IEEE. Secondly, the example of CEN-CENELEC shows that, although civil society groups are not yet as present as the private sector in AI standardisation, they are more widely included insofar as AI has a strong impact on the fundamental rights of individuals and standards grow increasingly closer to normative issues. Thirdly, the production of recent AI standards, such as ISO/IEC 42001, and numerous initiatives within CEN-CENELEC JTC 21, show that the trend towards quality management standards – or more generally “governance” standards – is far from over. Like ISO 9001 in its day, 42001 could become the world’s most widely adopted standard for standardising AI production in companies. But this trend is not limited to ISO, as the National Institute of Standards and Technology (NIST) in the US is also tending to focus more on open frameworks, such as the NIST Risk Management Framework (RMF) for AI ([Choi, 2024](#)). Finally, we note with AI standards that the standardisation process based on consensus is predominant at both ISO/IEC and CEN-CENELEC, and that the discourse these organisations hold is a unifying one, eager to display a high degree of legitimacy. Yet, they remain highly political and diplomatic spaces, rather than pure technical ones.

These four tendencies are not independent from one another, as we witness a diversification of standard setters in general, whether in terms of organisation types or individual participation. Similarly, the diversification of profiles, including towards less technical profiles, has accelerated the proliferation of standards that are also less technical, and has forced processes to become, at least in appearance, more democratic, in order to listen to this plurality of voices.

¹⁴[Yates and Murphy \(2019\)](#) mostly mentions point (i) and (iii). The two other points are very complementary but should, in our opinion, be separated from the former two.

AI standardisation goes further than the third wave

However, while AI standardisation appears to tick the boxes of the third wave of standardisation, it also has some distinctive elements which lead us to believe that a fourth wave of standardisation could be on the horizon.

To begin with, despite the very dominant position of international standards organisations such as ISO and transnational professional organisations such as IEEE, national and regional standards initiatives are also gaining in visibility. This is the case with CEN-CENELEC, which is seeking to break free from the influence of international standards and bring a more European touch to standards by developing standards more respectful of “European values”. In addition, national institutes, such as the NIST in the US, are developing their AI frameworks independently of ISO and other international bodies, include them in local policies and laws¹⁵, while gaining in visibility outside their home country. But this trend towards local standards is unlikely to replace international standards as these two types of initiatives serve different objectives. We therefore believe that a return to the establishment of national or regional standards, as in the first wave of standardisation, is highly unlikely. However, we could see a proliferation of AI standards, with many different entities producing their own framework. This was already the case with the AI ethics charters which, in addition to company charters, have also been drawn up by various government institutions in different countries or regions, such as the HLEG in Europe, and by supranational organisations such as the OECD¹⁶ or UNESCO¹⁷. Like AI ethics charters, this trend where each entity develops its own AI standards could start with companies. We are already seeing the emergence of private AI standards from a single company, such as Microsoft’s (Microsoft, 2022). If this trend towards the diversification of sources continues with future AI standards, we will have the first element to consider it as a new wave of standardisation.

If non-technical profiles such as civil society groups are more included with AI, we also see a diversification of the “technical” profiles, where company experts are less experts in AI itself and more experts in the governance of standardisation processes. This is mainly due, at least in the context of AI, to the inclusion of experts from BigTech companies, who aim to disseminate corporate objectives in the development of standards. To this end, they take on various roles within these standards bodies, getting involved in initiatives related to AI, but also in a variety of other areas. This diversification of profiles towards “governance” experts is the second elements of this fourth wave of standardisation.

Additionally to the proliferation of quality management standards, we are seeing the emergence of new areas of interest, focusing more on normative issues, such as standards for ethics or fundamental rights. If the first examples of standards on normative issues goes back to GDPR and its standards on privacy and data management, and to the IEEE standards on the ethics of robotics (Bryson and Winfield, 2017), it takes on a new dimension with AI, particularly in Europe where CEN-CENELEC is actively working on these subjects. But Europe is not alone in working on standards relating to normative issues. IEEE, for instance, is continuing to publish new standards on the ethics of autonomous systems, and ISO/IEC is also tackling the social issues associated with AI. More specific standards linked to ethical principles combined with algorithmic practices are also being developed, such as fairness standards. This new normative dimension is distinctive of AI standards. Although technical choices have always been value-laden, the topics that AI standards choose to address are more political than before and raise questions about their practical

¹⁵NIST standards are cited in official communication of the US Department of State ([Bureau of Cyberspace and Digital Policy, 2024](#)).

¹⁶Organisation for Economic Co-operation and Development.

¹⁷United Nations Educational, Scientific and Cultural Organization.

implementation.

Finally, while the processes for developing AI standards are based on consensus¹⁸ and all interest groups in society are called upon to participate, this apparent democracy is hindered by the stronger participation of large companies. While this was already known in the literature covering the third wave of standardisation (Mattli and Buthe, 2003; Brunsson and Jacobsson, 2002; Büthe and Mattli, 2011), with AI, traditional sector-specific corporations are being replaced by BigTech companies, with horizontal views on AI, which are deliberately keeping discussions at a higher level. In addition, these players have understood the power of this structure based on national bodies and have infiltrated these different national groups to ensure their dominant position overall in the SDOs. This crushing of other stakeholders seems to be characteristic of AI standards, or at least more prevalent than in other areas, due to the gigantic economic power held in particular by the BigTech, as a mirror of their position in our modern society. To fight against this stranglehold, in Europe, the Commission could decide to set aside the ESOs and to develop its own standards: the “common specifications” for AI (art. 41 AI Act). These specifications will probably not be developed using traditional consensus-based processes, but rather the Commission could choose who develop them, thus hijacking these deliberative processes. With AI standards could therefore come the end of consensus-based standardisation processes, for which we saw the limits in the current power dynamics within SDOs.

4 The obstacles faced by the European standardisation system force the Commission to intervene

AI standards inherit a system that was not designed to support legislation or to deal with a technology with so many social implications. As a result, it may be ill-suited to deal with AI-specific issues. In this situation, what levers for action does the Commission have?

4.1 A lack of trust in standardisation

The European standardisation system faces a number of obstacles today, such as the length of its processes, the business model of selling standards even when they are harmonised, the adoption of international standards to support European legislation and the stranglehold of large companies on the content of standards. Senden (2017) therefore questions the “*constitutional fit*” of the European standardisation system, as there do not appear to be sufficient safeguards with regard to the delegation of power, competence and implementation. There is therefore a lack of trust in standardisation, including from the European Commission.

A first indicator of this lack of trust in the standardisation system is the gradual integration of common specifications into EU regulations, including the AI Act. These common specifications come alongside a long list of various means of compliance created by the AI Act, with different levels of recognition: codes of practices, codes of conducts, delegated acts, implementing acts, and guidelines. The multiplication of these frameworks shows European Commission’s desire to have more control on the technical content of the AI Act. With the exception of codes of conduct and guidelines, all documents will be superseded by hENs, but only if hENs exist and meet the Commission’s expectations. The variety of documents planned therefore reveals the need for safety nets in case hENs fail.

The lack of trust in standardisation is also reflected in the sidelining of ETSI from the AI Act’s standardisation request. [P2] recounts: “[*ETSI’s*] governance is not very well accepted by

¹⁸At least in traditional SDOs such as ISO/IEC or CEN-CENELEC.

the European community. They are extremely efficient, but at the same time, in ETSI, whoever pays more has more power”*. The Commission therefore fears that CEN-CENELEC will become like ETSI: very open to international voices and controlled by the stakeholders who have the most money – i.e. experts – to invest.

Although CEN-CENELEC is far from being in the same situation as ETSI, its relationship with the Commission is also shifting, with EU officials becoming increasingly involved in standardisation and exercising stricter control over hENs in the standards development phase, whereas previously they were only involved in validating the end of the harmonisation process with the publication in the OJEU. The Commission notably tries to communicate as best it can with the ESOs on the types of deliverables they want: standards that are adapted to EU legislation and sufficiently precise (Soler Garrido et al., 2024), but with a diverse range of voices including academics, SMEs and civil society organisations in addition to traditional industry.

Even after the standards have been developed, the Commission could still strengthen its control over the harmonisation process. Indeed, some experts believe that the Commission can only base its assessment on a strict comparison between the scope of hENs and the scope of the corresponding EU regulation and standardisation request (Ebers, 2022). Yet, the ecosystem newly created by the AI Act, including the AI Office and its 140 employees – among which are technology specialists – could help bolster the Commission’s manpower for reviewing standards.

4.2 A mismatch between EU demands and what is possible

AI standards are expected to be both technically precise, but also address all AI systems at the same time, while encompassing normative issues such as the protection of fundamental rights, and complementing the law. They are the cornerstone of the European strategy for AI regulation, but are unable to meet expectations because of how much they are required to do and the complexity of the task at hand. This raises the question of whether we are asking too much of standards. There is a mismatch between the possible outcome of standardisation, i.e. what the system can do, and what is expected by Europe.

One example of this mismatch is the promotion by EU institutions of a more diverse standardisation system, with more academics and civil society representatives, while initially standardisation is understood as a consensus between private actors. This demand goes hand in hand with the desire to integrate ethical, social and fundamental rights issues into standardisation and legitimise the ESOs.

But this diversification is not always well received by standardisation experts. A study from the United Nations notes that there is in standardisation a *“prevalent perception that including human rights considerations would hinder efficient, speedy standard development and implementation processes, as it would require building new expertise and the participation of more actors”* (OHCHR, 2023). We have also witnessed this discourse, with experts, even in humanities, believing that civil society stakeholders *“follow an agenda that is highly non-technical”* and which *“slows down the process [of standardisation]”* [P16].

Yet in the context of AI standards as requested by the Commission, i.e. dealing with fundamental rights issues, the involvement of these stakeholders is essential. In addition, the participation of academics allows standardisation groups to have more experts in computer science without depending on industry to decide what should be considered the state of the art. We believe that all these perspectives are needed to develop standards that reflect not just what the industry *wants* to do, but also what it *should* do.

4.3 A variety of possible improvements

Rethinking the system?

This mismatch between the EU’s needs and the solutions that the ESOs can provide raises the question of alternatives to this standardisation system. We could imagine a system where the Commission recreates its own standardisation bodies. In the past, the Commission has considered the idea of creating an independent “European Agency for Standards” to manage the standardisation process in place of the ESOs (European Commission, 2011). This would enable the Commission to have a closer relationship with the standards body, avoid the rejection of standardisation requests and have standards closer to the Commission’s concerns. However, the Commission acknowledged that this could only be done with the support of the NSBs and that it would not necessarily reduce development times. Finally, the idea was not pursued for cost reasons, with the Commission stating that this solution “*would certainly lead to substantial additional costs for the EU-budget*”¹⁹. Indeed, in addition to the structural upheaval it would create, it will probably not be sufficient to make this new body legitimate. A simpler solution would be to improve the current system brick by brick.

On the business model of standards

But even if the current standardisation system is experiencing some difficulties for AI, there is still room for improvements. First of all, the business model for standards has recently been called into questions by case laws such as the Public.Resource.Org case. Although the decision from the Court of Justice of the European Union (CJEU) does not entail the systematic publication of hENs free of charge, it may set a precedent for more open publications. However, the issue of the copyright on standards has yet to be addressed, as it may conflict with the right of access to legal acts such as hENs. The business model of hENs needs to be further clarified, particularly with regard to whether or not it is suitable to support European law. In addition, the SDOs are currently thinking about improving their products and services, with “smart” standards that could be implemented directly²⁰. The emergence of this new type of standard could revolutionise the standards business model and may solve the issue of right of access. Nevertheless, we recommend improving access to these documents as they have strong legislative power and can have a direct impact on citizens’ rights, even more so where standards, as in the case of the AI Act, actively attempt to address normative issues such as fundamental rights.

There also appears to be a funding problem, with the Commission stating in the early 2010s that it contributes to almost half of the total income of the three ESOs, even though hENs account for a small portion of their deliverables (European Commission, 2011)²¹. The Commission could decide to make this funding conditional on free access to standards. However, such conditions would mean that the Commission and the ESOs recognise that their relationship is similar to outsourcing, with the standardisation request being similar to a service contract.

On participation in standardisation

While the European Commission considers that “*meaningful participation*” (European Commission, 2011) in standardisation requires a baseline of approximately 20% of a person’s time, this is

¹⁹The additional cost is estimated to be at least €20,000,000.

²⁰According to Baeva et al. (2023), smart standards are machine readable standards that could be automatically evaluated and verified.

²¹11.9% of CEN-CENELEC standards are cited in the OJEU today (CEN-CENELEC, 2024). It should be noted, however, that these two studies were carried out more than ten years apart, but we were unable to find more recent figures.

currently much higher in AI standardisation, as there are very few people actively involved compared to the task in hand. Indeed, while standards are necessary in our modern society, standardisation is an activity invisible to most people (Spivak and Brenner, 1993), citizens, scholars and small companies alike. Solutions to this lack of interest include launching awareness campaigns, teaching the role of standardisation at school – in particular in engineering and law programmes, providing additional funding to enable experts to participate, and so on. To improve effective participation and maintain people’s interest in standardisation once they are registered, training could be provided, for example by NSBs.

A key stake is the participation of European SMEs in the ESOs. SMEs find it difficult to benefit from and become involved in standardisation, as they come up against a number of obstacles, such as a lack of awareness of the existence and importance of standards, access to or understanding of standards, and the cost of investment (de Vries et al., 2009). Several solutions were proposed by the standardisation ecosystem and scholars for the better inclusion of SMEs. (de Vries et al., 2009) proposes the development of national strategies for the education on standardisation, the reinforcement of communication efforts specific to these stakeholders, or the possibility to reduce costs to buy standards. More recently, for AI standardisation, the working group 1 of CEN-CENELEC JTC 21 created a Task Group dedicated to “inclusiveness”, in charge of raising awareness about JTC 21 activities and bring new stakeholders to the discussions (CEN-CENELEC, n.d.a).

Since standards are to tackle technologies that have a high impact on citizens rights, the involvement of civil society stakeholders is also key to the preservation of these rights. However, despite the Commission’s calls for more diversity, the participation of these stakeholders is still scarce. Ebers (2022) proposes that societal stakeholders should be granted voting rights in the development of international standards, rights of appeal, and unlimited access to technical bodies and advisory groups, as well as to existing standards, without charge. Galvagna (2023) proposes to amend Regulation 1025/2012 on standardisation (European Parliament and Council, 2012) to include more participants in Annex III who could be funded by the Commission and participate in the work at the ESOs, to fund more individuals from civil society organisations with the StandICT system, and to create or fund a central hub to support civil society participation.

Finally, to include more academics, solutions include greater funding, including from StandICT, as well as recognition for individual work on standards. Universities and research centres could also encourage their researchers to take part in standardisation by offering them specific job opportunities and put in place communication campaigns to highlight the importance of their work. In general, a better understanding of standardisation by the academic system could improve the recognition and valorisation of standardisation activity.

On transparency

The legitimacy problems encountered by ESOs are mainly due to their opacity and lack of transparency, both internally and externally. Indeed, the ESOs do not usually communicate on who their experts are or what they are working on (Perarnaud, 2023). People working in the field of standardisation also point out that it is sometimes difficult to know which company a certain expert works for, as it is not compulsory to disclose this information and, on a voluntary basis, it is not always easy to get people to comply and many refuse to [P5]. Indeed, experts are supposed to represent their own opinion, not that of their company or entity of origin. In practice, however, it is difficult to distinguish between the two.

Measures could be taken to make the standards development process more transparent. If minutes are already taken, they sometimes do not contain all the underlying information and debates. We therefore suggest that meetings be recorded and that the recordings be made available to par-

ticipants in the standardisation process. Affiliations should be made known at standardisation meetings to enable experts to understand the position of their colleagues and to identify lobbying activities. As far as external communication is concerned, we have already suggested that the names of the main contributors should appear on the standards, particularly in the case of academics. If this solution is rejected, an alternative might be to reveal only the names of companies taking part in standardisation work, not necessarily the number of experts they have, but at least the fact that they are involved and on which projects.

All these solutions could help to improve the current system without revolutionising its structure. They do not require major investment and should not be too complicated to implement.

4.4 Deadline for standards push the ESOs to consider international standards

Many of the experts we interviewed criticised the lengthy processes of standardisation which hamper the drafting of standards. Today, stakeholders admit that drafting standards, either harmonised or not, take about five years [P5]. But that does not mean that after five years, the standardisation working groups will necessarily produce a standard. Sometimes, processes get blocked.

One of the solution put forward by the Commission in the early 2010s to reduce development time, was to simply put a strict deadline for the ESOs to draft hENs in the hope that processes will be accelerated. In return, the Commission committed to deliver standardisation requests to the ESOs more quickly (European Commission, 2011). However, today, we see that this strategy has proven inefficient, at least for AI standards. The Commission delivered the request before the final publication of the AI Act, which enabled CEN-CENELEC to start working on standards early on. At the same time, the request provided for a deadline in early 2025, only three years after the request was accepted by the ESOs. However, the experts point out that having to work with both a legal text under construction and a standardisation request dating from the first proposal caused some conflicts, as it was not clear which version of the text or request the ESOs should refer to [P5]. Furthermore, the deadline imposed by the Commission did not encourage the ESOs to improve their internal processes, but only gave stakeholders who were opposed to the development of new standards more arguments to push for the adoption of international standards.

The adoption of international standards instead of the drafting of new European standards in the context of AI was pushed for early on. Before the standardisation request and the forming of an official JTC on AI, CEN-CENELEC experts on AI were grouped in a Focus group. In a response to the White paper on AI, the focus group explains the role that standards can play in an upcoming European regulation (CEN-CENELEC, 2020). They notably advocated for the adoption of ISO/IEC foundational standards as a reference for upcoming European standards. They also advocate for the consideration of ISO risk management standards for the risk-based approach. This desire to rely on international standards was also displayed in the Focus group Roadmap (CEN-CENELEC, 2020). As confirmed by the experts we interviewed, some stakeholders, such as BigTech companies, supported the creation of a European JTC that would be an “*empty shell*”*, simply adopting existing international standards without carrying out any additional work [P5]. These arguments were backed up by the fact that international standards were already published and that, if European work was to start from scratch, it would take a lot of man power, lead to overlaps and inconsistencies with international standards and standards might not even be delivered in time for the implementation of the AI Act. The adoption of international standards as hENs is, however, not new. Indeed, 28% of all hEN from CEN are adapted from ISO standards, and 69% for CENELEC and IEC (CEN-CENELEC, 2024). For instance, the medical device regulation, often cited by EU officials as the most convincing example of the NLF’s success, has almost all its harmonised

standards derived from ISO²². With the support of the European Commission, experts opposed to BigTech managed to get the JTC to reach a consensus on the inadequacy of international standards alone to cover the AI Act. However, there are still major disagreements today that slow down standardisation work, including on the adoption of international standards for risk management.

The example of the AI Act therefore shows that, today, the question of the timing between standards and European legislation is not entirely resolved. However, we do not believe that the Commission could have done things differently. There was a rush for standards as there was a rush for the AI Act because the technology is evolving fast and regulation must try to keep up. Ideally, for future regulation, the Commission should try to give more time to standardisation experts, or give them the means, in terms of staff and money, to develop standards more calmly.

4.5 If standards are not ready in time, the Commission will have to make a strong political choice

The AI Act places so much emphasis on technical standards that it is possible to take them for granted and forget that they do not yet exist (Pouget, 2023). The Commission’s deadline of April 2025 is considered at best ambitious, at worst unrealistic (Perarnaud, 2023). Several voices explain that it is still unsure when the hENs will be ready, although it will be, in any case, too late for the initial deadline. Experts now tend to think that a postponement to 2026 is inevitable. Nevertheless, standards must be ready several months before the provisions relating to high-risk AI systems come into force in August 2026, to give companies time to comply (Soler Garrido et al., 2024). However, this obvious difficulty on the part of CEN-CENELEC in delivering standards on time is at odds with the reassuring tone of the European Commission. In 2021, the Commission announced “*a large set of relevant harmonised standards could be available within 3-4 years*” (European Commission, 2021) and, in a recent webinar, EU officials remained positive about the ability of the ESOs to meet the deadlines (European Commission, 2024).

But if the standards are indeed not ready by April 2025, or even 2026, which now seems likely, the Commission will have to decide what to do. A first option could be to publish parts of the hENs in the OJEU, showing what the ESOs have achieved by the deadline. This would maintain the *status quo*, but runs the risk of creating loopholes in standards that providers could exploit given the legal power of hENs. Further work would help fill these gaps, but they would remain for a certain period of time. A second solution would be to wait for a complete version of all hENs. However, this could take years as some items, such as cybersecurity requirements, are far from ready [P16]. Meanwhile, alternative frameworks, which the Commission does not control and which may come from the industry or any other entity worldwide, could emerge and be used by companies to assess their conformity. Furthermore, there is no guarantee that the standards which are being developed will satisfy the Commission and pass the HAS assessment. hENs may therefore never be published in the OJEU and never be granted legal power. This situation, in which a hENs has been offered for citation in the OJEU but was refused, has already occurred in other fields, but remains fairly rare²³. Finally, the Commission could choose to develop common specifications to fill the gaps during the wait for hENs. However, it is not clear exactly where the Commission would find the manpower and how quickly this could be achieved. The legitimacy of common specifications, which would be drafted “*on the back of an envelope*”*, hijacking the consensus-based standardisation processes, could also be compromised.

There are two questions that the Commission must consider. First, what will be the impact on

²²A list of these standards can be found at: (European Commission, n.d.c).

²³6.5% of CEN-CENELEC standards which were offered for citation are currently not cited, either because the process is underway or because it was refused (CEN-CENELEC, 2024).

the implementation of the AI Act? Indeed, further economic and legal studies will have to confirm whether the publication of parts of standards is better than nothing, or whether the legal power of hENs is too important to be left to semi-finished work. Second, what will be the impact on the EU’s legislative framework and its legitimacy? Indeed, by waiting past the deadline, by not publishing hENs at all, or by choosing to draft its own specifications, the Commission would be acknowledging the failure of hENs and of the European standardisation system as a whole. At a time when all eyes are on the Commission to monitor the implementation of the newly adopted legal text, such bad press could have consequences far beyond AI, as it could be seen as a sign that the current standardisation system is not, or is no longer, fit to support EU legislation. The decision that the Commission will have to take is therefore highly political and will reveal the extent to which it still believes in the current system.

5 Fixing AI standards by refocusing on vertical sectors

5.1 European AI standards are horizontal because they follow the AI Act

When we spoke with standardisation experts, they explained that AI standards are very high-level today and rarely go into technical details, such as mathematical definitions, metrics or thresholds. They rather remain fairly general, with definitions in natural language. We felt that academics and computer scientist experts in particular were frustrated by the lack of technical content. The experts identify several reasons for this general approach to standards, depending on the type of stakeholder to which they belong. Industry experts usually argue that the methods and metrics are not sufficiently mature, and that including them in a standard would be taking a risk: the standard might not work in practice, for example, or might quickly be rendered obsolete by a method that was recently discovered. In general, these industry experts are more focused on implementation. On the contrary, academics tend to believe that industry experts do not want standards to be too technically precise, because less stringent requirements would allow them to be less constrained by regulation. These academic experts are therefore more concerned with ensuring that standards are close to the technical state of the art in research.

However, there are also other reasons why standards remain very general. First of all, AI standards are still in their infancy, with the first to be developed being ISO/IEC standards from 2018 onwards. As standardisation work has to start somewhere, it is only natural to have broader standards at the beginning and more specific requirements later on.

Second of all, the term “AI” is itself quite broad and covers a wide range of contexts, both in terms of (i) technology, and (ii) sector of application. Indeed, AI can be a simple machine learning model such as a classifier, a rule-based model, a deep neural network used for natural language processing or another network with a different structure used for image recognition, a reinforcement learning model, and so on. The same image recognition technology, for example, could be deployed in healthcare to detect cancer, in an autonomous vehicle to analyse road signs, or simply to carry out an internet search. These different contexts imply different ethical stakes and the requirements for AI systems are likely to vary depending on the technology and the sector of use.

However, despite these contextual differences, the AI Act itself is very horizontal. Although it lays down different requirements for AI systems depending on their risk category, the categorisation of systems is not strictly based on technology or sectors. Systems can always fall into a certain category depending on their context. AI systems used for education or justice, for example, are immediately classified as high-risk because these sectors are included in Annex III. Systems used in the healthcare sector will also be classified as high-risk if they fall under the medical devices regulation listed in Annex II. Similarly, some technologies fall directly into one risk category. For

instance, biometric systems are automatically classified as high risk and, in certain circumstances, can also be classified as unacceptable risk. However, although risk categories are sometimes constructed by listing different technologies or sectors, the requirements applicable to them are neither technology-specific nor sector-specific. All high-risk AI systems, whether they are used for education or healthcare, whether they use natural language processing or image recognition, will have to comply with the same requirements in terms of data quality, transparency, accuracy and so on. This horizontality is specific to the AI Act, as other EU safety regulations are product-specific (Ebers, 2022). One of the experts we interviewed, [P16], sums up the situation as follows: *“the major challenge we are facing [in standardisation] is that the AI Act is a product safety regulation approach. And AI is not a product. AI is a component of products”*.

Policy documents prior to the AI Act show that the European Commission’s approach with the AI Act was first and foremost to avoid the duplication of sectoral legislation that would each impose their own requirements in an inconsistent manner. To remain in line with the AI Act’s approach of having the same requirements for all high-risk AI systems, regardless of their technology or sector of use, the standards that will be harmonised and support the legislation must also be horizontal. They must therefore address in a general manner each of the requirements of the AI Act – transparency, accuracy, etc. This horizontality is even cited by the European Commission as a desirable characteristic of AI standards (Soler Garrido et al., 2024). If CEN-CENELEC had proposed a standard by technology or by use case, this would probably not have passed the check of the HAS consultant, who assesses in particular whether the standards being harmonised are in line with their respective legislation.

5.2 Standards need context

In addition to its horizontal nature, the AI Act draws heavily on ethical charters and policy documents that are based on the concept of “trustworthiness”, i.e. a list of different ethical and technical properties required of AI systems. However, these documents are usually quite broad, simply defining principles or ethical-technical properties in general. On the contrary, a part of these documents focuses on very technical elements, sometimes by limiting the field of study to one technology – such as facial recognition, or one sector – such as education. The conclusion seems to be the same for ethics charters as for standards: it is possible either to stay at the level of AI in general and define broad properties without going too much into technical detail, or to dive deeper into a use case and define more technical elements, give more precise definitions, metrics and methods more aligned with the context of use.

For instance, the performance of an AI system is calculated very differently depending on whether it is a classifier or a language model. In the case of a classifier, a simple measure of accuracy or f1-score might suffice. For language models, many metrics might exist depending on which task the model is trying to perform. Indeed, models can focus on processing tasks such as sentiment analysis, text classification, inference or semantic understanding. They could also generate, translate or summarise text. Each of these tasks will therefore require a specific metric (Chang et al., 2024). Similarly, fairness metrics are different in regular classification models, and in facial recognition models for instance. While classification tasks usually use group fairness measures such as demographic parity or equalised odds, fairness metrics in facial recognition are often based on a comparison between the error rates of the two groups on which the system performs best and worst, or on the average of all error rates (Grother, 2022).

In some cases, some trustworthiness criteria might not even be relevant. Fairness, for instance, is more important in recruitment than in industrial production (Baeva et al., 2023) In recruitment, ensuring fairness involves ensuring that performance is similar between groups of people, for example

between men and women. In this case, fairness and good practice in mitigating bias will have a direct impact on fundamental rights. On the other hand, group fairness in industrial production only concerns the correct functioning of the algorithm for different types of construction parts. It will therefore have no direct impact on individuals.

5.3 Other harmonised standards are vertical

The AI Act is one of the first product safety regulations adopted under the NLF not to be strictly limited to a single sector. For comparison, other regulations, such as for toys or medical devices, refer to a particular industry: the toy industry or the healthcare industry. The standards referenced in the OJEU are therefore also vertical in these sectors. This verticality of previous standards is not a coincidence: according to [Choi \(2024\)](#), standards work best for narrow tasks for which they can give quantitative attributes. This is particularly true of software standards, which have become less effective since they have been reduced to large open-ended frameworks. [Choi \(2024\)](#) therefore calls for standards to be refocused on more specific tasks in order to make them more effective, and for very broad quality management standards that leave too much flexibility to industry to be banned.

The only exception to this verticality of NLF regulations is Regulation 2024/1781 on ecodesign requirements for sustainable products ([European Parliament and Council, 2024b](#)) – formerly Directive 2009/125/EC ([European Parliament and Council, 2009](#)). The 2024 Regulation is too recent to have hENs, and there are no references to hENs published in the OJEU under the 2009 Directive. Indeed, the former Ecodesign Directive, as well as the new Ecodesign Regulation, are what European institutions call “*framework legislation*”. In the words of the [European Commission \(n.d.b\)](#), this means that “*concrete product rules will be decided progressively over time, on a product-by-product basis, or horizontally, on the basis of groups of products with similar characteristics*”. These concrete product rules are published in the form of Commission’s implementing measures, such as for computers, dishwashers, ovens, fans, washing machines, etc²⁴. As a result, although the regulation is horizontal, the hENs for ecodesign requirements are sector-specific on these products. A few standards retain this horizontal approach to enable the alignment between vertical standards, but they are not harmonised. However, this raises the problem of the interplay between horizontal and vertical standards, as well as a lack of clarity surrounding the role of horizontal standards ([Bundgaard and Huulgaard, 2023](#)). Horizontal ecodesign standards can serve as a “*shopping list*” and ensure alignment ([Bundgaard and Huulgaard, 2023](#)), but discussions are still ongoing to see if they can be used directly for certain products or if product standards should be developed first ([Schischke et al., 2022](#)).

5.4 Developing vertical standards will help with various issues

Refocusing AI standards on vertical sectors could help in several ways. First, it would help solve practical implementation issues by defining technical requirements that are feasible in a certain context if used. It would also support the regulation much more efficiently by declining each legal requirement into verifiable quantifiable criteria, which are impossible to define today at the horizontal level. The result would be more an effective regulation that AI system providers would be obliged to comply with, rather than interpreting in their own way, which could lead to abuse.

Verticality could also be a solution to the integration of values in standards as certain values could be prioritised in a given context²⁵. Indeed, the hierarchy between principles is never addressed

²⁴A list of these standards can be found here: ([European Commission, n.d.a](#)).

²⁵[Gonzalez Torres and Ali-Vehmas \(2024\)](#) give the example of AI for voting which could prioritise the value of

in ethics charters, nor is the hierarchy between requirements in the AI Act. If in theory, all legal requirements should be respected – as well as all ethical principles – in practice, some requirements might be working against each other, such as fairness and accuracy. Contextual technical requirement might help address how a compromise could work for a given technology in a given sector.

The development of vertical standards could also help to solve some of the structural problems facing the European standardisation ecosystem. More vertical and therefore more technical standards could attract more academics and researchers, who are more interested in the technical details of standards. This would also help to involve SMEs, which are generally more interested in the practical aspects of implementing standards in their sector than in the general standards that large companies are currently pushing at European level [P13]. More vertical standards would therefore attract domain specialists, complementing the knowledge of AI experts who may not necessarily have in mind the necessary details for the proper deployment of an AI system in a given sector. Ultimately, this would provide a counterpower to the voice of larger companies, and a greater willingness to publish standards.

5.5 Horizontal standards are not necessarily a first step towards vertical ones

As AI standards are still in their infancy, it is difficult to predict what they will look like in the future. Even if standardisation efforts are horizontal today, some others vertical ones may develop in the future. Some experts therefore remain convinced that vertical standards will come after the publication of horizontal standards [P15]. However, defining vertical standards after horizontal standards still raises practical difficulties.

Firstly, having given general elements and definitions could prevent vertical standards from defining their own terms and requirements. As alignment is crucial in standardisation, if horizontal standards define terms – such as accuracy or transparency – in a certain way, it is virtually impossible for other subsequent standards, such as vertical standards, to go against this definition. However, the same term may have different definitions in different contexts. If the definitions are decided at the level of AI in general, they may not work for certain technologies and certain contexts. Secondly, even if vertical standards develop at a later stage, the harmonised standards that will prevail and have legal effect will always be the less precise horizontal standards.

5.6 Is it too late to change the strategy?

Standardisation experts believe that it is too late to change the ESOs standardisation strategy and opt for a vertical approach to AI standardisation [P16]. Indeed, CEN-CENELEC JTC 21 has been working for several years on the horizontal “AI Trustworthiness framework”, neglecting vertical ones. A few exceptions are projects on technology-specific standards, such as Natural Language Processing (NLP) or computer vision. However, other technologies and sectors have not yet been addressed. While some groups at international level are dealing with systems that would fall under the definition of AI, such as ISO/IEC JTC 1/SC 37 on biometrics, these initiatives are also still in their infancy and European working groups seem to be struggling to develop. Ideally, there should be one working group per sector affected by the AI Act, including Annex II on product safety regulation – e.g. an AI working group for medical devices, one for AI in machinery, one for AI in toys, etc – as well as Annex III on sectors of application – e.g. a working group for AI in education, one for AI in critical infrastructure, etc. In addition to working groups on high-risk AI

democracy.

systems, we could consider working groups on GPAI models separately in anticipation of a potential standardisation request on this topic.

However, such a high number of working groups and standards are simply too high for the ESOs to handle. Even when we discussed the AI Trustworthiness framework with experts, they told us that, at this pace, it will be difficult to pass the HAS assessment so that it could be published in the OJEU [P16]. Not all the items that should be covered will be ready in time, and the framework might not correspond to what the Commission had in mind. We therefore propose that the AI Trustworthiness framework should be adopted by CEN-CENELEC when it is ready, to be a starting point for other standards, but it should not be harmonised. Once this work is finished, JTC 21 should start working on vertical standards that are technology-specific and sector-specific. However, these vertical standards cannot be harmonised since the standardisation request was for horizontal standards on AI and the Commission has insisted on the need for horizontal hENs (Soler Garrido et al., 2024). This solution entails that there will be no hENs published in the OJEU for the AI Act, therefore no standards that gives a presumption of conformity to the law. But if the vertical standards are adopted, there will still be standards that companies can use for conformity assessment.

6 Conclusion

The AI Act was not the beginning of AI regulation at the European level. Indeed, the work of the HLEG started well before and already sets the goals of the European ecosystem: reaching for all high-risk AI systems, a certain level of “trustworthiness”, understood as a the respect for a list of technical properties. The AI Act continues on this notion of trustworthiness by proposing obligations that high-risk systems will have to respect, that are based on technical properties which will be defined in standards, and more precisely in harmonised standards. As of today, the work of CEN-CENELEC JTC 21 focuses mostly on the “AI Trustworthiness framework”, a broad horizontal standard meant to tackle all of the essential requirements for high-risk AI systems. However, this broadness and horizontality, inherited from the AI Act and the work of the HLEG, prevents its from defining specific technical requirements which are technology- and context-specific. This is notably true for standards which incorporate ethical and fundamental rights elements, a trend more and more prominent in AI standardisation. In this position paper, we propose that the three spheres of “trustworthy AI”, i.e. ethics, law and technique, should be kept separate. We note that AI standardisation might mark the beginning of a fourth wave of standardisation, where standards deal with increasingly normative questions, though we believe that experts should remain cautious when dealing with these issues. The European standardisation system could be improved and we provide some possible solutions for more transparent processes and a more diverse participation. However, CEN-CENELEC will need to accelerate its process of standard setting if it wants to deliver on time. If hENs are not up to the taste of the Commission, they could be rejected during the assessment process and not have their references published in the OJEU. This strong political choice will have to be made by the Commission, but it would then be acknowledging the partial failure of the current standardisation system. We propose that, in this case, future standards should focus on vertical sectors, where they could go much further into technical detail rather than remaining at an unsatisfactory high level. Given that the successful implementation of the AI Act will probably depend on the effectiveness of its relevant technical standards, the choice of whether or not to reference hENs in the OJEU will be crucial.

References

- Baeva, D. G., Puntschuh, M., and Binder, M. (2023). Power to the standards: Expert consultation on the role of norms and standards in the European regulation of artificial intelligence (White paper). The Center for Trustworthy Artificial Intelligence. <https://www.zvki.de/zvki-exklusiv/fachinformationen/zvki-whitepaper-power-to-the-standards>
- Bjørnskov, C. (2017). Social Trust and Economic Growth. In Uslaner, E. M., editor, *The Oxford Handbook of Social and Political Trust*, pages 535–556. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190274801.013.24>
- Bolte, L., Vandemeulebroucke, T., and van Wynsberghe, A. (2022). From an Ethics of Carefulness to an Ethics of Desirability: Going Beyond Current Ethics Approaches to Sustainable AI. *Sustainability*, 14(8):4472. <https://doi.org/10.3390/su14084472>
- Brunsson, N. and Jacobsson, B. (2002). *A World of Standards*. Oxford University Press. <https://global.oup.com/academic/product/a-world-of-standards-9780199256952>
- Bryson, J. and Winfield, A. (2017). Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. *Computer*, 50(5):116–119. IEEE. <https://ieeexplore.ieee.org/document/7924235>
- Bundgaard, A. M. and Huulgaard, R. D. (2023). The role of standards in support of material efficiency requirements under the Ecodesign Directive. *Journal of Cleaner Production*, 385(135599). <https://doi.org/10.1016/j.jclepro.2022.135599>
- Bureau of Cyberspace and Digital Policy (2024). Risk Management Profile for AI and Human Rights. U.S. Department of State. <https://www.state.gov/risk-management-profile-for-ai-and-human-rights/>
- Busch, L. (2011). *Standards: Recipes for Reality*. The MIT Press. <https://doi.org/10.7551/mitpress/8962.001.0001>
- Büthe, T. and Mattli, W. (2011). *The new global rulers: the privatization of regulation in the world economy*. Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691144795/the-new-global-rulers>
- Cath-Speth, C. J. N. (2021). *Changing minds and machines: a case study of human rights advocacy in the Internet Engineering Task Force (IETF)*. PhD thesis, University of Oxford. <https://ora.ox.ac.uk/objects/uuid:9b844ffb-d5bb-4388-bb2f-305ddedb8939>
- CEN-CENELEC (2018). Code of Conduct for delegates, experts and observers participating in the technical work of CEN and/or CENELEC. <https://boss.cen.eu/reference-material/Guidancedoc/Pages/ExpertsCodeconduct>
- CEN-CENELEC (2020). CEN-CENELEC response to the EC White Paper on AI. https://www.cencenelec.eu/media/CEN-CENELEC/Areas%20of%20Work/Position%20Paper/cen-clc_ai_fg_white-paper-response_final-version_june-2020.pdf
- CEN-CENELEC (2020). Road Map on Artificial Intelligence (AI). Technical report, CEN-CENELEC Focus Group on AI. https://www.standict.eu/sites/default/files/2021-03/CEN-CLC_FGR_RoadMapAI.pdf

- CEN-CENELEC (2024). CEN CENELEC in figures – 2024 Q3. https://www.cencenelec.eu/stats/CEN_CENELEC_in_figures_quarter.htm
- CEN-CENELEC (2024). Status Dashboards JTC21 European AI standardisation. https://www.linkedin.com/posts/sebastianhallensleben_status-dashboard-jtc21-european-ai-standardisation-activity-7235665875112988673-xfEH
- CEN-CENELEC (n.d.a). Artificial Intelligence. <https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/>
- CEN-CENELEC (n.d.b). CEN/CLC/JTC 21 Published Standards. https://standards.cencenelec.eu/dyn/www/f?p=205:32:0::::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D
- CEN-CENELEC (n.d.c). CEN/CLC/JTC 21 Work programme. https://standards.cencenelec.eu/dyn/www/f?p=205:22:0::::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., and Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, 15(3):1–45. <https://doi.org/10.1145/3641289>
- Choi, B. H. (2024). NIST’s software un-standards. *Lawfare*. <https://www.lawfaremedia.org/article/nist-s-software-un-standards>
- de Vries, H., Blind, K., Mangelsdorf, A., Verheul, H., and van der Zwan, J. (2009). SME access to European standardization. Rotterdam School of Management, Erasmus University. https://www.erim.eur.nl/fileadmin/default/content/erim/content_area/news/2009/smeaccessreport%202009.pdf
- DIN and DKE (2022). German Standardization Roadmap on Artificial Intelligence (2nd Edition). <http://www.din.de/go/roadmap-ai>
- Ebers, M. (2022). Standardizing AI - The Case of the European Commission’s Proposal for an Artificial Intelligence Act. In DiMatteo, L. A., Poncibò, C., and Cannarsa, M., editors, *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*. Cambridge University Press. <https://doi.org/10.1017/9781009072168.030>
- EDRi (2022). The role of standards and standardisation processes in the EU’s Artificial Intelligence (AI) Act. *European Digital Rights (EDRi)*. <https://edri.org/wp-content/uploads/2022/05/The-role-of-standards-and-standardisation-processes-in-the-EUs-Artificial-Intelligence-AI-Act.pdf>
- Eliantonio, M. and Medzmariashvili, M. (2017). Hybridity Under Scrutiny: How European Standardization Shakes the Foundations of EU Constitutional and Internal Market Law. *Legal Issues of Economic Integration*, 44(4). <https://doi.org/10.54648/leie2017017>
- European Commission (2011). Commission Staff Working Paper – Impact Assessment Accompanying document to the Proposal for a Regulation of the European Parliament and of the Council on European Standardisation and amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC,

2009/105/EC and 2009/23/EC. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011SC0671>

European Commission (2021). Commission Staff Working Document – Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. SWD/2021/84 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084>

European Commission (2023). Commission implementing decision on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence (C(2023)3215). [https://ec.europa.eu/transparency/documents-register/detail?ref=C\(2023\)3215&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en)

European Commission (2024). 1st European AI Office webinar on Risk management logic of the AI Act and related standards. Shaping Europe’s digital future. <https://digital-strategy.ec.europa.eu/en/events/1st-european-ai-office-webinar-risk-management-logic-ai-act-and-related-standards>

European Commission (n.d.a). Ecodesign and Energy Labelling. Internal Market, Industry, Entrepreneurship and SMEs. https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards/ecodesign_en

European Commission (n.d.b). Ecodesign for Sustainable Products Regulation. https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/ecodesign-sustainable-products-regulation_en

European Commission (n.d.c). Harmonised standards. Public Health. https://health.ec.europa.eu/medical-devices-topics-interest/harmonised-standards_en

European Parliament (2010). European Parliament resolution of 21 October 2010 on the future of European standardisation (2010/2051(INI)). https://www.europarl.europa.eu/doceo/document/TA-7-2010-0384_EN.html

European Parliament (2023). Report on a standardisation strategy for the single market. A9-0136/2023. Committee on the Internal Market and Consumer Protection. Rapporteur: Bielan, A. https://www.europarl.europa.eu/doceo/document/A-9-2023-0136_EN.html

European Parliament and Council (2009). Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products (recast). OJ L 285, p.10-35. <http://data.europa.eu/eli/dir/2009/125/oj/eng>

European Parliament and Council (2012). Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/eec and 93/15/eec and Directives 94/9/ec, 94/25/ec, 95/16/ec, 97/23/ec, 98/34/ec, 2004/22/ec, 2007/23/ec, 2009/23/ec and 2009/105/ec of the European Parliament and of the Council and repealing Council Decision 87/95/eec and Decision no 1673/2006/ec of the European Parliament and of the Council. OJ L 316, p.12–33. <https://eur-lex.europa.eu/eli/reg/2012/1025/oj>

- European Parliament and Council (2024a). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). OJ L. <http://data.europa.eu/eli/reg/2024/1689/oj>
- European Parliament and Council (2024b). Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of eco-design requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/EC. Legislative Body: EP, CONSIL. <http://data.europa.eu/eli/reg/2024/1781/oj/eng>
- Galvagna, C. (2023). Inclusive AI governance. Technical report, Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/>
- Giovannini, C. (2021). The role of standards in meeting consumer needs and expectations of AI in the European Commission proposal for an Artificial Intelligence Act. Technical report, The European consumer voice in standardisation (ANEC). <https://www.anec.eu/images/Publications/position-papers/Digital/ANEC-DIGITAL-2021-G-141.pdf>
- Gonzalez Torres, A. P. and Ali-Vehmas, T. (2024). Governing Through Standards: Artificial Intelligence and Values. Delft University of Technology. <https://easychair.org/publications/reprint/xgpk>
- Gornet, M. and Herman, H. (2024). A peek into European standards making for AI: between geopolitical and economic interests.
- Gornet, M. and Maxwell, W. (2023). Normes techniques et éthique de l’IA. In *Conférence Nationale en Intelligence Artificielle (CNIA @ PflA 2023)*, Strasbourg, France. Association Française pour l’Intelligence Artificielle (AFIA). <https://pfla23.icube.unistra.fr/conferences/cnia/publications/Gornet.pdf>
- Gornet, M. and Maxwell, W. (2024). The European approach to regulating AI through technical standards. *Internet Policy Review*, 13(3). <https://doi.org/10.14763/2024.3.1784>
- Grother, P. (2022). Face Recognition Vendor Test (FRVT) Part 8: Summarizing Demographic Differentials. Technical Report NIST IR 8429, National Institute of Standards and Technology (NIST). <https://doi.org/10.6028/NIST.IR.8429.ipd>
- Groves, L., Metcalf, J., Kennedy, A., Vecchione, B., and Strait, A. (2024). Auditing Work: Exploring the New York City algorithmic bias audit regime. In *FACCT ’24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1107–1120. Association for Computing Machinery. <https://doi.org/10.1145/3630106.3658959>
- Hagendorff, T. (2020). The ethics of ai ethics: An evaluation of guidelines. *Minds and machines*, 30(1):99–120. Springer. <https://doi.org/10.1007/s11023-020-09517-8>
- Harding, S. (1992). After the Neutrality Ideal: Science, Politics, and “Strong Objectivity”. *Social Research*, 59(3):567–587. The Johns Hopkins University Press. <https://www.jstor.org/stable/40970706>

- HLEG (2019). Ethics guidelines for trustworthy AI. Technical report, Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. Publications Office of the European Union, Directorate-General for Communications Networks, Content and Technology. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Hunyadi, M. (2015). *La Tyrannie Des Modes De Vie. Sur Le Paradoxe Moral De Notre Temps*. Le Bord de l'Eau. <https://www.editionsbdl.com/produit/la-tyrannie-des-modes-de-vie-sur-le-paradoxe-moral-de-notre-temps/>
- Hunyadi, M. (2018). Which Ethics for Bioethics? The Example of CRISPR-Cas9. Institut protestant de Paris, Faculté de Théologie. <http://hdl.handle.net/2078.1/196134>
- ISO/IEC (2020). ISO/IEC TR 24028:2020, Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. Edition 1. <https://www.iso.org/standard/77608.html>
- ISO/IEC (2023). ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system. Edition 1. <https://www.iso.org/standard/81230.html>
- Laux, J., Wachter, S., and Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1):3–32. John Wiley & Sons. <https://doi.org/10.1111/rego.12512>
- Laux, J., Wachter, S., and Mittelstadt, B. (2024). Three pathways for standardisation and ethical disclosure by default under the European union artificial intelligence act. *Computer Law & Security Review*, 53. Elsevier Advanced Technology. <https://doi.org/10.1016/j.clsr.2024.105957>
- Loya, T. A. and Boli, J. (1999). Standardization in the World Polity: Technical Rationality Over Power. In Boli, J. and Thomas, G. M., editors, *Constructing world culture : international nongovernmental organizations since 1875*. Stanford University Press. <https://www.sup.org/books/title/?id=2040>
- Mattli, W. and Buthe, T. (2003). Setting International Standards: Technological Rationality or Primacy of Power? *World Politics*, 56(1):1–42. Johns Hopkins University Press. <https://doi.org/10.1353/wp.2004.0006>
- Merriam-Webster Dictionary (n.d.). Ethics. <https://www.merriam-webster.com/dictionary/ethics>
- Microsoft (2022). Microsoft Responsible AI Standard, v2. General Requirements. <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf>
- Nickel, P. J., Franssen, M., and Kroes, P. (2010). Can We Make Sense of the Notion of Trustworthy Technology? *Knowledge, Technology & Policy*, 23(3):429–444. <https://doi.org/10.1007/s12130-010-9124-6>
- OHCHR (2023). Human rights and technical standard-setting processes for new and emerging digital technologies : Report of the Office of the United Nations High Commissioner for Human Rights. Office of the United Nations High Commissioner for Human Rights. <https://digitallibrary.un.org/record/4031373>

- Perarnaud, C. (2023). With the AI Act, we need to mind the standards gap. *Centre for European Policy Studies*. <https://www.ceps.eu/with-the-ai-act-we-need-to-mind-the-standards-gap/>
- Pouget, H. (2023). The EU’s AI Act Is Barreling Toward AI Standards That Do Not Exist. *Lawfare*. <https://www.lawfaremedia.org/article/eus-ai-act-barreling-toward-ai-standards-do-not-exist>
- Probasco, E., Toney, A., and Curlee, K. (2023). The Inigo Montoya Problem for Trustworthy AI. Technical report, Center for Security and Emerging Technology. <https://doi.org/10.51593/20230014a>
- Rieder, G., Simon, J., and Wong, P.-H. (2021). Mapping the Stony Road toward Trustworthy AI: Expectations, Problems, Conundrums. In Pelillo, M. and Scantamburlo, T., editors, *Machines We Trust: Perspectives on Dependable AI*. MIT Press Direct. <https://doi.org/10.7551/mitpress/12186.003.0007>
- Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, 26(5):2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>
- Saltelli, A., Dankel, D. J., Di Fiore, M., Holland, N., and Pigeon, M. (2022). Science, the endless frontier of regulatory capture. *Futures*, 135. Pergamon. <https://doi.org/10.1016/j.future.s.2021.102860>
- Schischke, K., Berwald, A., Dimitrova, G., Rückschloss, J., Nissen, N. F., and Schneider-Ramelow, M. (2022). Durability, reparability and recyclability: Applying material efficiency standards EN 4555x to mobile phones and tablet computers. *Procedia CIRP*, 105:619–624. <https://doi.org/10.1016/j.procir.2022.02.103>
- Senden, L. (2017). The Constitutional Fit of European Standardization Put to the Test. *Legal Issues of Economic Integration*, 44(4). <https://doi.org/10.54648/leie2017018>
- Soler Garrido, J., De Nigris, S., Bassani, E., Sanchez, I., Evas, T., André, A.-A., and Boulangé, T. (2024). Harmonised Standards for the European AI Act. *JRC Publications Repository*. <https://publications.jrc.ec.europa.eu/repository/handle/JRC139430>
- Spivak, S. M. and Brenner, F. C. (1993). *Standardization Essentials: Principles and Practice*. CRC Press. <https://doi.org/10.1201/9781482277388>
- Tartaro, A. (2024). Value-laden challenges for technical standards supporting regulation in the field of AI. *Ethics and Information Technology*, 26(4):72. <https://doi.org/10.1007/s10676-024-09809-y>
- Tessier, C. (2022). Parler du numérique et de son éthique: un questionnement... éthique. In *Pour une éthique du numérique*, pages 97–105. Presses Universitaires de France (PUF). <https://www.puf.com/pour-une-ethique-du-numerique>
- Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, 2(2):46–48. Nature Publishing Group. <https://doi.org/10.1038/s41928-019-0213-6>
- WTO (2000). Principles for the Development of International Standards, Guides and Recommendations. World Trade Organization. https://www.wto.org/english/tratop_e/tbt_e/principles_standards_tbt_e.htm

Yates, J. and Murphy, C. N. (2019). *Engineering Rules: Global Standard Setting since 1880*. Johns Hopkins University Press. <https://muse.jhu.edu/pub/1/monograph/book/66187>