Kenniscentrum Data & Maatschappij

ETHICAL PRINCIPLES AND (NON-)EXISTING LEGAL RULES FOR AI

Knowledge Centre Data & Society

October 2021

Citation: Knowledge Centre Data & Society, "Ethical principles and (non-)existing legal rules for AI", October 2021

For contact: jan.debruyne@kuleuven.be

www.data-en-maatschappij.ai

GENERAL INFORMATION

Context of the guide - The European Commission (EC) based its AI strategy on three pillars: (i) investment in research to advance the capacity and use of AI, (ii) preparation for socio-economic change, and (iii) developing an appropriate ethical and legal framework consistent with European Union (EU) values.

With regard to point (iii), the work of the High-Level Expert Group on Artificial Intelligence (AI HLEG), founded in June 2018, is crucial. In April 2019, the AI HLEG published the Ethics Guidelines for Trustworthy AI. This document contains a number of recommendations for the development of trustworthy AI. Trustworthy AI systems comply with the law, act in accordance with ethical values/principles and are robust (both from a social and technical perspective). The Guidelines contain seven ethical requirements that AI systems need to meet in order to be trustworthy:

- Human agency and oversight;
- Technical robustness and safety;
- Privacy and data governance;
- Transparency;
- Diversity, non-discrimination and fairness;
- Societal and environmental well-being; and
- Accountability.

Assessment List for Trustworthy AI (ALTAI) - These Ethics Guidelines are not applicable law. They merely give guidance on how AI can be ethically designed, developed and used (soft law). Nevertheless, it is recommended to take these key requirements into account as much as possible to ensure that AI systems are developed in accordance with these principles. However, how to comply with these Ethics Guidelines is not always clear and/or straightforward. An Assessment List for Trustworthy AI (ALTAI) was therefore added to the Ethics Guidelines. This list contains a number of specific questions to assess compliance with a given ethical requirement.

Structure of the guide - In this guide, the Knowledge Centre Data and Society (KDS) examines the extent to which the questions in the ALTAI are already reflected in the existing legal framework and whether there are certain relevant rules that are consistent with the ethical requirements. We also highlight the instances in which there is still room for clarification or refinement. We explore each of the seven ethical requirements in more detail via the following questions:

- What does the ethical requirement mean?
- Which rules already reflect the ethical requirement or can serve as inspiration for adopting (additional) provisions in addition to the requirement?
- Where are the possible points of improvement and/or points of attention?
- What tools are already in place/could be used to fulfil the ethical requirement?

Aim of the guide - The content of this guide is presented in an interactive webpage, i.e. in a kind of accessible online encyclopaedia (in Dutch). This guide serves as a basis/framework for additional actions (cf. methodology that was used for the KDS guide on AI and the GDPR and corresponding fact sheets). The webpage that is launched by the KDS follows the structure of this guide, with a subdivision for each ethical requirement (first layer) with corresponding drop-down options underneath each question as outlined above (second layer).

The guide and webpages serve three objectives. Firstly, we intend to provide an overview of the applicable ethical requirements with a brief substantive framework and interpretation. In this regard, we also identify the relevant legislation. We do not analyse this legislation in detail but highlight its essence and integrate hyperlinks to where it can be found. Secondly, for each ethical requirement, we aim to identify a number of regulatory gaps and formulate several recommendations for policy makers. On the basis of this, more targeted actions can be taken

in the future. Thirdly, the intention is to develop a working document in which the information will be updated at regular moments and for which stakeholders can always give feedback or recommendations.

Conception of the guide - This guide was developed through consultation with and input from stakeholders and with the support of the Flemish Department of Economy, Science & Innovation (EWI). Following an internal consultation, a general overview of the ethical requirements was drawn up. This was given to the stakeholders for feedback. Researchers at the KU Leuven Centre for IT & IP Law (CiTiP) are responsible for coordinating this guide. The stakeholders involved and the actors involved in the Flemish Policy Plan on AI were also given the opportunity to provide feedback on a draft version of this guide. Additional feedback, remarks, questions and input on/about this guide can be provided at any time to researchers working at CiTiP.

About the KDS - The Knowledge Centre Data and Society is a collaboration between three university research groups: imec-SMIT-VUB, KU Leuven CiTiP and imec-MICT-UGent. It is part of the Flemish Policy Plan on Artificial Intelligence and receives support from the Flemish government (EWI). The KDS is the central hub for the legal, societal and ethical aspects of data-driven and AI applications. The KDS wishes to contribute to the debate and garner public support for AI and data-driven applications. The information provided by the KDS, including this guide, is general and cannot be regarded as individual legal advice. It cannot be a substitute for advice from a legal expert. While the KDS strives to ensure that documents are correct and accurate, the position contained therein may not be applicable to your particular situation, may not be complete, accurate or current, or may not reflect the position that a court or regulatory authority may take. The KDS therefore assumes no responsibility whatsoever for compliance with applicable legal requirements on the part of an organisation.

ETHICAL REQUIREMENT 1: HUMAN AGENCY AND OVERSIGHT

What does the ethical requirement mean?

This first ethical requirement refers to the fact that AI systems should support **human agency** and **human decision-making**, as prescribed by the principle of respect for human autonomy. This requires that AI systems should act as enablers for a democratic, flourishing and equitable society by supporting the user's agency and uphold fundamental rights, which should be underpinned by human oversight.

Under this requirement, AI systems are evaluated on the basis of **two sub-components**, namely the extent to which they respect (1) human agency and (2) allow human oversight.

Human agency addresses the effect that AI systems can have on human behaviour in the broadest sense. It deals with the effect of AI systems that are aimed at guiding, influencing or supporting humans in decision-making processes, for example algorithmic decision support systems or risk analysis/prediction systems (i.e. recommender systems, predictive policing, financial risk analysis). It also deals with the effect on human perceptions and expectations when confronted with AI systems that 'act' like humans. Finally, it deals with the effect of AI systems on human affection, trust and (in)dependence.

The requirement of **human oversight** helps to self-assess necessary oversight measures. This oversight can be achieved through governance measures such as human-in-the-loop (HITL), human-on-the-loop (HOTL) or human-in-command (HIC) approaches.

- Human-in-the-loop refers to the capability for human intervention in every decision cycle of the system.
- Human-on-the-loop refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation.
- Human-in-command refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the AI system in any particular situation. The latter can include the decision not to use an AI system in a particular situation to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by an AI system.

Keep in mind that, depending on the scope of application of the AI system and the potential risks, **more or less far-reaching oversight measures** may be needed to support other safety and control measures. If all other conditions remain the same, an AI system should be tested more extensively and stricter governance measures may be required when less human oversight of the system becomes possible.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides several questions for each element evoked. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration.

1) ***Human autonomy***

	Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society? ● Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision? ● Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision? Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system? ● Are end-users or subjects informed that they are interacting with an AI system?

➔ Data Protection

Under Art. 13 and 14 of the General Data Protection Regulation (GDPR), controllers must provide certain information to the individuals whose personal data they are processing. If there is a situation of **automated decision-making or profiling**, this must be acknowledged, and meaningful information must be provided about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

The mirror image of this is Art. 22 of the GDPR, which obliges controllers to **provide** data subjects with at least the **right** to **obtain human intervention** on the part of the controller, to express his/her point of view and to contest the decision. In order for data subjects to be actually and sufficiently able to challenge such an automated decision, it can be required that they are informed that the decision was generated by an AI system and/or that they interacted with such a system.

More information can also be found in the section on Ethical Requirement 4: Transparency.

➔ Consumer protection

There are also **information obligations** in consumer protection law, which can already result in the sellers of AI systems being obliged to inform end users that a decision or outcome is the result of an algorithmic decision, or that they are interacting with an AI system.

For example, under the general information obligation in Art. VI.2 of the Code of Economic Law (CEL), consumers must be informed of the **main characteristics** of the products, where appropriate, the **functionality** of digital content (including applicable technical protection measures) and the relevant **interoperability** of digital content with hardware and software and other services of which the company is aware or can reasonably be expected to be aware. Similar provisions also extend to services (Art. VI.45 and VI.64 CEL) and financial services (Art. VI.55 CEL).

➔ General contract law

If an end user or consumer is not adequately informed of the fact that a decision or outcome is the result of an algorithmic decision or that they are interacting with an AI system which is used in a situation where a contract

is concluded, this may lead to **error/mistake** (*dwaling*) on the part of the user or consumer (Art. 1110 Civil Code, note that the numbering will be replaced once the New Book on Obligations will become applicable).

Error is a defect of consent which, in principle, prevents the formation of a valid contract. In essence, error means that a party to a contract has misrepresented reality at the time the contract was concluded. Furthermore, there needs to be a 'substantial error', which means that the contract would not have been concluded without this error. As a simple illustration, suppose someone orders an artwork from an online service selling artworks by a well-known artist (e.g. Rinus Van De Velde), where in fact these artworks are actually generated by an AI system. If the buyer trusts that these artworks are produced by the artist him/herself and, therefore, orders a painting, he/she will be able to invoke error if he/she subsequently finds out that the painting was created by an AI system.

Such an error must nonetheless be 'excusable'. This requirement implies that a reasonable person placed in similar circumstances would also have made the same error. Applied to the illustration above, this means that the buyer could not have inferred from the information provided by the service provider, nor on the basis of any investigation he/she made himself, that the artworks were AI creations and not authentic work.

	Could the AI system affect human autonomy by generating over-reliance by end-users? ● Did you put procedures in place to avoid that end-users over-rely on the AI system? Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way? ● Did you put in place procedures to avoid that the AI system inadvertently affects human autonomy?

➔ Consumer protection

In this context, we can highlight the fact that consumer law prohibits so-called **unfair misleading actions** (Art. VI.97-100 CEL). A commercial practice shall be regarded as misleading if it contains false information and is therefore untruthful or if it in any way, including through its overall presentation, deceives or is likely to deceive the average consumer, even if the information is factually correct, in relation to given elements, and in either case causes or is likely to cause him/her to take a transactional decision that he would not have taken otherwise. The elements on which misleading information can be provided include:

- the existence or nature of the product;
- the main characteristics of the product (including its risks, composition, fitness for purpose, usage, etc.)

One can also speak of **misleading omissions**. This is the case if essential information which the average consumer needs, according to the context, to take an informed transactional decision is omitted and which causes or is likely to cause the average consumer to take a transactional decision that they would not have taken otherwise. This includes situations in which the essential information is provided in an unclear, unintelligible, ambiguous or untimely manner, or if the commercial intent, if not already apparent from the context, was not made clear and, in either case, the average consumer would or would be likely to take a transactional decision that he/she would not have taken otherwise.

As such, if an AI system (e.g. chatbot) is used to sell products or services, it must be ensured that the system would not (start to) use misleading information, but always provide a consumer with the correct information or

at least refer to it. If this is not the case, an end user's decision-making process could be disrupted in unintended or undesirable ways.

➔ General contract law

Although it may be a little far-fetched, organisations using an AI system in the course of contract formation must also ensure that the AI system does not involve 'physical force' or 'coercion'/'duress' (Art. 1111-1115 of the Civil Code, note that the numbering will be replaced once the new Book on Obligations will be applicable). **Coercion**, like error, is a vitiated consent and presents an obstacle to a valid contract. In practice, this does not often occur. More specifically, coercion consists of exercising a physical or moral constraint, or at least the threat thereof, with respect to the co-contractor him/herself, his/her honour or his/her property.

It therefore needs to be ensured that an AI system selling specific products does not go too far and (starts to) exert undue coercion on potential buyers to perform its function. This could for example be the case if the system threatened to hack into the buyer's digital devices or those of his/her family. Ensuring that AI does not exert undue coercion avoids any disruption to the end-user's decision-making process in unintended or undesirable ways.

➔ Criminal law

In this context, we can also highlight the risk that an AI system would perpetrate or be used in the context of possible criminal **offences** such as computer fraud, extortion, scamming and deception or abuse of trust. These are further discussed in the section Ethical Requirement 2: Technical Robustness and Safety.

	Does the AI system simulate social interaction with or between end-users or subjects? Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour? Depending on which risks are possible or likely, you can answer the questions below: ● Did you take measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System? ● Did you take measures to minimise the risk of addiction? ● Did you take measures to mitigate the risk of manipulation?

➔ Consumer protection

Regarding the risk of disproportionate attachment and manipulation, the provisions on **misleading and aggressive commercial practices** are relevant.

For example, Art. VI.100 CEL stipulates that it is a **misleading commercial practice** to claim that products facilitate winning at games of chance or, in the context of a commercial practice, to claim that a competition is being organised or prizes awarded without actually awarding the prizes described or a reasonable alternative. As such, an AI system that sells products or services must not make such claims and it must be avoided that the system begins making such claims while performing its function.

Furthermore, Art. VI.103 CEL stipulates that it is an **aggressive commercial practice** for a business to persistently contact a consumer unsolicited by telephone, fax, e-mail or other remote media. This may also include a risk of manipulation. It must therefore be avoided that AI systems perform such actions.

➔ General contract law

Manipulation, in the form of deception, is also illegal under general contract law. **Deception** is a third variant of vitiated consent, and presents an obstacle to concluding a valid contract (Art. 1116 of the Civil Code, note that the numbering will be replaced once the new Book on Obligations will be applicable).

For there to be a situation of deceit, a party must use **duplicity**, in bad faith, in order to deliberately mislead the other party into concluding a contract. An error must therefore be deliberately elicited without which the contract would not have been concluded. In other words, the duplicity must have been decisive for the conclusion of the contract. This duplicity can be positive acts (e.g. lying), providing false/incomplete information or deceptive silence or concealment of information.

➔ Gambling legislation

In this context, we can also consider the **stake limit** imposed by Belgian gambling legislation in certain situations, not only limiting the gambling loss, but also intending to reduce the risk of addiction. As such, if games incorporating AI are introduced, it must also be ensured that they respect any applicable stake limits.

2) Human oversight

	Did you determine whether the AI system (choose as many as appropriate): ● is a self-learning or autonomous system; ● is overseen by a Human-in-the-Loop; ● is overseen by a Human-on-the-Loop; ● is overseen by a Human-in-Command. Did you ensure that the humans (human-in-the-loop, human-on-the-loop, human-in-command) have been given a specific training on how to exercise oversight?

➔ Employment Law

In this context, we can in the first place think of employers who use AI systems in their production processes or services. Indeed, on the one hand they have to ensure the safety of their employees in the workplace. On the other hand, they have to allow them to exercise their profession, among other things by giving instructions. It is therefore important that they give employees the right and appropriate training if they come into contact with AI systems.

More specifically, Art. 4 of the Act regarding well-being at work stipulates that the well-being of employees must be promoted by taking measures relating to, inter alia, workplace safety and the psychosocial aspects of the work. Art. 5 and 6 go into more detail in this regard.

Art. 5 for example lists some of the prevention principles that an employer must apply. These include **preventing risks**, reducing the risks of serious injury by taking material measures, **informing** workers about the nature of their work, the risks to which it gives rise and the associated measures, and providing workers with **appropriate instructions** and guidance to reasonably ensure compliance with these instructions. Conversely, Art. 6 stipulates that every worker must take care of his/her own safety and health and that of other relevant persons to the best of his/her ability and in accordance with his/her training and the instructions given by the employer.

More information can also be found in the section Ethical Requirement 6: Environmental and Societal Well-being.

➔ General and specific product safety rules

In this context, we can also highlight the obligation of producers to **provide instructions or information** in the context of product safety. For example, Art. IX.8 CEL stipulates that producers must provide users with information on the risks inherent in a product throughout the normal or reasonably foreseeable period of its use, where such risks are not immediately obvious without adequate warnings.

The obligation to provide information and instructions is also addressed in **sectoral legislation**. For example, toy safety legislation requires manufacturers to accompany the toys they make with instructions and safety information. Information obligations also exist for medical devices.

Taking into account the self-learning nature of AI products, such a provision would once again appear to be relevant. More information can be found in the section on Ethical Requirement 4: Transparency.

	Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject? Did you ensure a 'stop button' or a procedure to safely abort an operation when needed? Did you take specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?

➔ Data Protection

The GDPR requires the data controller to conduct a **data protection impact assessment** (DPIA) in certain cases. A DPIA must be carried out prior to any processing likely to present a high risk to data subjects. It is thus a specific obligation to give prior consideration to the risks that the processing of personal data may pose to the rights and freedoms of natural persons (Art. 35).

For a detailed discussion of this, also see Ethical Requirement 3: Privacy and Data Governance.

➔ Safety at work

Also in this context, we can highlight the fact that the employer must promote the **well-being of employees** by taking measures relating to, inter alia, workplace safety and the psychosocial aspects of the work. Indeed, an employer needs to apply certain prevention principles. As such, employers using AI systems in their production processes or services can establish similar warning and response methodologies, 'stop button' procedures or control measures to comply with this obligation.

More information can also be found in the section Ethical Requirement 6: Environmental and Societal Well-being.

➔ Safety standards

There are various safety standards which already impose specific **risk assessments** and require products to meet **specific minimum health and safety requirements**. Those provisions may therefore be relevant in this context.

One can for example think of the **Machinery Directive** that was transposed by the Royal Decree of 12 August 2008. The Machinery Directive has a very specific scope (Art. 1). It stipulates that manufacturers of machinery must carry out a risk assessment to determine the health and safety requirements that apply to the machinery. The results of this risk assessment must then be taken into account in the design and construction of the machinery. This also applies to the control systems (e.g. AI software) of machinery which, among other things,

need to be developed in such a way that errors in the control system logic do not lead to hazardous situations. Moreover, the Machinery Directive in certain cases also requires that each machine must be fitted with one or more **emergency stop devices** to enable actual or impending danger to be averted.

The **European Low Voltage Directive**, which is transposed by the Royal Decree of 21 April 2016, imposes on manufacturers of electrical equipment the requirement to draw up technical documentation including a risk analysis and assessment. This assessment needs to take into account the general requirement that electrical equipment can only be placed on the EU market if, when correctly installed and maintained and used for its intended purpose, it does not endanger the health and safety of people, domestic animals or property. For example, such electrical equipment must, among other things, be provided with protection against dangers that may arise from external influences.

Other specific safety rules that may be relevant in this context are those applicable to **toys**.

More generally, there are also the **product safety rules** contained in Book IX CEL. For example, Art. IX.8 CEL stipulates that producers (possibly supported by distributors) must take measures to remain informed of the risks associated with their products and/or services, and take appropriate action to prevent those risks. These measures include an indication of the identity and contact details of the producer on the product or packaging. This will allow consumers to lodge complaints or perform random checks on products placed on the market.

Also see Art. 14 of the Proposal of an EU Regulation on AI on human oversight on high-risk AI systems.

Where are the possible points of improvement or focus points?

Clarification (applicability) of concepts

The current regulations were not designed for systems that can 'learn' or take unforeseen actions by the designer/user. Indeed, it is generally assumed that the company or natural person who uses certain tools has control over them and therefore final responsibility. Autonomous AI systems are at odds with this tacit assumption.

Consequently, it would appear important for policy makers to **clarify** the extent to which designers and users of AI systems **need to take measures or envisage procedures with the aim of avoiding undesirable consequences** due to the self-learning/autonomous nature of AI systems.

On the other hand, these questions also relate to less clear-cut notions such as **'addiction'** and **'attachment'**, which are not yet covered in current law. The question here is whether, and to what extent, the (European or Belgian) legislator can or should act to **convert these concepts into legal requirements**, not only for AI systems, but also for other products.

Safety standards

It is recommended to continue **developing and refining** safety standards related to AI products that take into account the self-learning nature of these products.

What tools can be used to fulfil the ethical requirement?

- Tarot Cards of Tech are designed to carry out an informal impact assessment during meetings through a brainstorming session.
- The Artificial Intelligence Impact Assessment (AIIA) is a structured method to clearly identify the (societal) benefits of an AI application. Moreover, there is also a focus on analysing the reliability, security and transparency of the AI system. The AIIA is a workshop that can be organised in the first stage of a project followed by regular feedback on the outcomes of the workshop.
- Data Ethics Decision Aid (DEDA) was created by the Utrecht Data School in cooperation with data analysts from the Municipality of Utrecht. It consists of various methods that can be used at the outset of a project. The aim is to document the ethical issues and the decisions made about them in order to be accountable to relevant stakeholders.
- The supporting ethics approach aims to enter into dialogue about the relationship between ethics and technology, in the form of a workshop, in order to gain a better overview of the possible ethical bottlenecks in the developed technology.
- The Ethical Explorer pack offers a free tool with a checklist in the form of a set of cards. This makes it possible to reflect on aspects in the development of ethical technology.
- The Product Impact Tool induces developers to think about the relationship between people and technology, and provides guidance on how to deal with new technologies.

ETHICAL REQUIREMENT 2: TECHNICAL ROBUSTNESS AND SAFETY

What does the ethical requirement mean?

The ethical requirement of technical robustness requires that AI systems are developed with a **preventative approach**. AI systems must behave reliably and as intended, while minimising unintentional and unexpected harm as well as preventing it where possible. The physical and mental integrity of people must always be safeguarded.

The requirement of technical robustness consists of **four sub-components**: (1) resilience to attack, and security; (2) fall back plan and general safety; (3) accuracy and (4) reliability and reproducibility.

AI systems must be **resilient to attacks and threats**. Vulnerabilities that allow AI systems to be attacked (e.g. by hacking) must be ruled out. This applies to both the software and the hardware in which the AI system is embedded. Indeed, such vulnerabilities can give rise to a wide range of damage.

In addition, for each AI system, a **fall back plan and general safety measures must be in place.** AI systems must perform their function without causing harm to people or the environment. Unintentional damage must also be avoided. Some form of risk assessment should also always take place with AI systems. The necessary level depends on the actual context: some AI systems entail high risk and require proactive testing. Other AI systems are low-risk and require less strenuous testing.

Furthermore, AI systems must be **accurate**. AI systems must be able to make correct considerations, for instance by classifying information into the right categories or making correct recommendations. That way, unintended risks of incorrect predictions can be reduced. AI systems must make it clear if errors cannot be avoided. Accuracy becomes even more important as the consequences of the AI system's decisions have an impact on the life or integrity of humans.

Finally, AI systems must be **reliable and reproducible**. Reliable AI systems work effectively with various inputs and in different situations. Reproducibility means that AI systems always work in the same way under the same conditions. As such, the behaviour of AI systems can be predicted. Replication files can simplify this work.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides several questions for each sub-component evoked. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration.

1) Resilience to attack and security

	Could the AI system have adversarial, critical or damaging effects (e.g. to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?

➔ Technical norms and standards

Most standards that are important for regulators are initially adopted by international (e.g. International Standards Organisation-ISO), European Union (e.g. CEN, CENELEC or ETSI) or national **standards bodies**. These are private organisations whose main activity is to produce standards for a variety of products (e.g. National Bureau for Standardisation).

These standardisation processes are still sometimes too slow to keep up with technological developments. For that reason, various new standards have been issued by industry consortia (e.g. World Wide Web Consortium (W3C) and Internet Engineering Task Force). The standards of the Institute of Electrical and Electronics Engineers (IEEE) are also important.

In principle, compliance with these standards is entirely voluntary, unless otherwise stipulated. In product safety legislation, compliance with the standards leads to a presumption of safety and compliance with the safety obligations. A **conformity assessment** is sometimes also required. These and other legislations are discussed below.

	Is the AI system certified for cybersecurity or is it compliant with specific security standards?

➔ Technical norms and standards such as the ISO standards on information technology and security techniques

The best-known cyber security standards are those of the **ISO 27000 family**, which prescribe tests related to the security management of information technology systems. Examples include ISO/IEC 27001:2017 and ISO/IEC 27002. At present, there are no harmonised EU standards for AI. At the ISO level and at some national levels, standards have been adopted for AI (e.g. ISO/IEC JTC 1/SC 42).

➔ Cybersecurity

Pursuant to the NIS Directive, the NIS Act provides a **presumption of security** for network and information systems certified in accordance with ISO 27001. Compliance with security requirements is demonstrated by a certificate issued by a competent accredited conformity assessment body.

Furthermore, the Cybersecurity Act provides for voluntary certification of ICT products, services and processes (Art. 52, 54 and 55). This certification framework is being prepared by the European Union Agency for Cybersecurity (ENISA) on behalf of the European Commission. A draft of a potential certification framework can be found at the following link.

➔ Data Protection

A detailed discussion on the GDPR is available in the explanation of Ethical Requirement 3 on privacy and data governance. All that needs to be stressed here is that the processing of personal data for **security purposes** may be carried out on the legal basis of legitimate interest (Art. 6(1)(f)).

The GDPR also provides that organisations can draw up **codes of conduct**. These may include agreements on how the processing is to be performed, how the rights of data subjects are to be exercised, what information is to be provided and what security measures are to be accepted (Art. 40). Art. 42 of the GDPR clarifies that **certification**

mechanisms are equally encouraged. Adherence to such codes of conduct constitutes proof of compliance with various obligations (see for example the relevant provisions in Art. 24 and 25).

➔ Product safety

Pursuant to the Product Safety Directive, Art. IX.2. CEL stipulates that the manufacturer may only place **safe products and services** on the market.

According to Art. IX.3 CEL, a product or service is **presumed to be safe** if it conforms to harmonised standards or (in the absence thereof) to applicable Belgian standards, recommendations of the European Commission, codes of conduct regarding product safety, the state of the art, the safety which a user may reasonably expect, and international standards.

➔ General safety obligation and specific safety rules

Recurring elements in the relevant regulations include the **presumption of safety** and the **safety obligation**. In addition, these rules also lay down specific safety rules for the **conformity assessment** of certain products (e.g. medical devices).

The forms of conformity assessment are described in Annex II of the EU Decision on a common framework for the marketing of products. In addition, these rules often require the affixing of the **CE marking** which proves that products comply with the product safety legislation. For certain products, an independent notified body has to be involved in the assessment of whether a product is in conformity and can therefore be CE marked. This can be found in rules per product category. The manufacturer must also draw up an EC declaration of conformity and the technical documentation. Examples include toy safety legislation or legislation on machinery. Another example is the legislation on electrical equipment.

A summary of the different categories of products and the applicable regulations and standards can be found at the following link.

➔ Common sales law, services law and contract law

Compliance with common contract law also ensures compliance with (safety) standards. This applies both to the purchase of goods (under Belgian law, this only applies to hardware or software installed on a carrier) and to the provision of services (this also applies to the provision of software without any carrier).

Art. 1604 and 1614 of the Civil Code provide that the seller must deliver a good to the buyer that is in conformity with the agreement ('**compliant delivery**'). This conformity also relates to the agreed quality. Service contracts also require the service provider to deliver the agreed or standard quality. It is therefore advisable to make clear agreements on this in a service level agreement.

Following Directive 1999/44, Art. 1649bis et seq. of the Civil Code provide for additional guarantees of conformity for sales of goods to **consumers**. The Vienna Sales Convention may also apply between companies based in different countries.

The new Directive 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services (Digital Content Directive) and Directive 2019/771 on certain aspects concerning contracts for the sale of goods (Consumer Sales Directive) both provide for an **obligation of conformity** for the supply of digital content (software, data, etc. supplied online) and consumer goods. This conformity is assessed both subjectively (based on the contract) and objectively (based on currently applicable standards and fitness for purposes of the product or the content). These Directives must be transposed into Belgian law by 1 July 2021 at the latest and will enter into force on 1 July 2022.

➔ Algorithmic trading

Following the MiFiD II Directive, the Act of 21 November 2017 on Infrastructures for markets in financial instruments provides for the obligation of market operators to ensure that effective systems, procedures and arrangements are in place to ensure that trading systems are resilient, have sufficient capacity to absorb volume peaks in orders and order messages, are able to ensure orderly trading under highly tense market conditions, are fully tested to ensure such conditions are met and are subject to effective business continuity arrangements to ensure continuity of the service provider in the event of a disruption to trading systems (Art. 22 §1 Law 21 November 2017).

This provision is repeated in the Royal Decree of 19 December 2017 laying down further rules for transposing the Markets in Financial Instruments Directive, which is referred to in the Banking Act and the Investment Firms Act.

Additional technical requirements for this obligation under MiFiD 2 are set out in Commission Delegated Regulation 2017/589 of 19 July 2016 with regard to regulatory technical standards specifying the organisational requirements of investment firms engaged in algorithmic trading.

How exposed is the AI system to cyber-attacks?

- Did you assess potential forms of attacks to which the AI system could be vulnerable?
- Did you consider different types of vulnerabilities to which the AI system could be vulnerable, such as data poisoning (i.e. manipulation of training data), model evasion (i.e. classifying the data according to the attacker's will) and model inversion (i.e. infer the model parameters)?

➔ Attacks

Like all other ICT systems, AI systems are susceptible to attacks.

Data poisoning entails that an attacker inserts new incorrectly labelled training data into the system to disrupt the behaviour of a self-learning system. AI systems can also be attacked via **model evasion techniques**. This involves exploiting vulnerabilities in the model so that it fails to recognise things or recognises them incorrectly. A final familiar tactic is **model inversion** in which the attacker tries to infer the underlying training data from a description of the AI model or its output.

➔ Data Protection

The controller is obliged to process data in such a way as to guarantee its **'appropriate' security**. Among other things, the data must be protected against unauthorised or unlawful processing and against accidental loss, destruction or damage (Art. 5(1)(f) of the GDPR).

Art. 32 GDPR also obliges the controller and the processor to implement appropriate **technical and organisational measures** to ensure a level of security appropriate to the risk. In certain cases, a **DPIA** must also be performed.

For more information, see Ethical Requirement 3: Privacy and data governance. By its very nature, a DPIA includes a modelling of potential threats.

➔ Practical methods and standards

In information security, **threat modelling** is used for risk analysis. This means that within a company, someone identifies all threats to the security of all information (including personal data) and takes the necessary measures. Familiar threat modelling tools include STRIDE (developed by engineers at Microsoft) and LINDDUN Privacy Threat Modelling (developed by KU Leuven). ISO 27001 also refers to a risk analysis and provides a list of risks that may be present for information security.

➔ Cybersecurity

The NIS Act stipulates that providers of essential services and digital service providers must take **appropriate and proportionate technical and organisational measures** to control the risks to the security of network and information systems on which their essential services depend. The provider must also take appropriate measures to prevent incidents affecting the security of network and information systems used for the provision of those essential services or to minimise their impact. To manage and minimise incidents, it must be possible to identify them, which equates to **threat modelling**.

The Cybersecurity Act stipulates that the purpose of the EU cybersecurity certification framework is, among other things, to **identify and document known dependencies and vulnerabilities** (Art. 51). It must be verified that ICT products, services and processes do not contain any known vulnerabilities and are secure by default and by design. There is once again a reference to a risk assessment, and therefore threat modelling.

➔ Algorithmic trading

The obligation to ensure resilience in the systems, including resilience against disruptions to the systems, as discussed above, can be seen as an obligation to be resilient against attacks from outside. The systems must also have been appropriately tested. Furthermore, Delegated Regulation 2017/589 also provides for the obligation of real time monitoring and an automated surveillance system to detect market manipulation.

➔ Possible sanctions

Various IT-related offences already exist, such as IT forgery (Art. 210bis Penal Code), IT fraud (Art. 504quater Penal Code), hacking (Art. 550bis Penal Code) and IT sabotage/unauthorised data manipulation (Art. 550ter Penal Code).

	Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?

➔ Cybersecurity

The **general obligations** of essential service providers and digital service providers under the NIS Act have already been discussed above. These measures apply throughout the lifecycle of the system. Provisions in the Cybersecurity Act are also relevant (e.g. Art. 51(j)).

➔ General Data Protection Regulation

The controller is obliged to **take appropriate measures** to ensure compliance with the GDPR. Among other things, appropriate measures include the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services. This implies constant security.

➔ Obligations of providers of electronic communications services

Following Art. 4 of the e-Privacy Directive, Art. 114 of the Act of 13 June 2005 on electronic communications requires providers of public electronic communication services to take appropriate technical and organisational measures to guarantee the security of their services. These should be proportionate to the risk involved thereby taking into account the state of the art and the cost of implementation.

➔ Consumer protection

The trader has to ensure that the consumer is informed of and supplied with updates, including security updates, that are necessary to keep the digital content or digital service in conformity (Art. 8 Digital Content Directive). The Consumer Sales Directive contains similar provisions for goods with digital elements (Art. 7).

➔ Product safety

Following the Product Safety Directive, Art. IX.8 §2 CEL stipulates that the producers of products and services must take measures commensurate with the characteristics of the products and services they supply in order to 1) keep informed of the risks of these products and services and 2) take the appropriate actions to avoid the risks of these products. Distributors also contribute to compliance with the safety requirements. Within the limits of their respective activities, they shall participate in monitoring the safety of products placed on the market (Art. IX.8 §3 CEL).

They also need to immediately inform the **Central Reporting Centre for Products** (*Centraal Meldpunt voor Producten*) when they know that the product or service poses risks to the user that are incompatible with the general safety obligation (Art. IX.8 §4 CEL). This obligation also applies to producers and distributors of products subject to specific safety legislation. Obligations are also sometimes imposed in other regulations (e.g. Art. 11 Construction Products Regulation).

➔ Algorithmic trading

The obligation to provide resilient systems for algorithmic trading has already been discussed above. Delegated Regulation 2017/589 provides, among other things, for annual self-assessment and real time monitoring to ensure continuous compliance with these requirements.

	Did you red-team/pen-test the AI system?[1]

➔ Criminal law

The illegality of hacking has already been covered. Regarding red team exercises or penetration tests to check the security of a network, it is recommended that there are **clear guidelines** about which IT systems can be infiltrated and how this can be done.

➔ Standards

The **methods** available for penetration tests include the Open Source Security Testing Methodology Manual, the OWASP Web Application Penetration Checklist, and the Penetration Testing Execution Standard.

[1] These are methods to test how well systems are protected against misuse.

➔ Regulation of investment firms engaged in algorithmic trading

The obligation to provide resilient systems has already been discussed for companies that use algorithms, for example to execute the buying and selling of securities.

Delegated Regulation 2017/589 provides for additional obligations. An investment firm must implement an **IT strategy** defining objectives and measures which is consistent with effective and secure IT management. An investment firm must perform **annual penetration tests** and vulnerability scans to simulate cyber-attacks (Art. 18).

	Did you inform end-users about the duration of security coverage and updates?

➔ Cybersecurity

In accordance with the Cybersecurity Act, the manufacturer and provider of ICT products, services and processes must, among other things, envisage the period **during which security support will be offered** to end users, in particular as regards the availability of cybersecurity related updates (Art. 55).

➔ Data Protection

The controller must inform the data subject, among other things, of the periods for which the data will be retained or the criteria used to determine the **period for which the personal data will be stored** (Art. 13(2) and 14(2) of the GDPR).

➔ Consumer protection

Within consumer law, there are various **information obligations** towards the end user. These are discussed under Ethical Requirement 4: Transparency.

➔ Contract law

The software vendor has an obligation to provide information about the main characteristics of the item in the event of a sale. The contractor must also perform his services in accordance with the rules of the art and has an information obligation as well. However, there is no legislation in this regard. For this reason, it is always advisable to have the duration of updates clearly stated in a **Service Level Agreement** in order to avoid any disputes.

	What length is the expected timeframe within which you provide security updates for the AI system?

➔ Consumer protection

The Digital Content Directive (Art. 8) and the Consumer Sales Directive (Art. 7) also contain provisions on **(security) updates**. The vendor will be obliged to provide all necessary security updates during the period envisaged in the contract that "the consumer can reasonably expect" in order to avoid liability. This also includes security updates.

➔ Product safety legislation

The **monitoring obligation** for producers regarding all risks related to their products has already been discussed (Art. IX.8 §2 CEL). These measures also include ensuring that producers of products containing software estimate the timeframe required for updates and that they provide updates throughout the life of the software. There is no provision for software producers.

2) *General safety*

	Did you define risks, risk metrics and risk levels of the AI system in each specific use case? Did you put in place a process to continuously measure and assess risks?

➔ General

The obligation to **conduct a risk analysis** of all threats and to carry out the necessary tests pursuant to the regulations on cybersecurity, data protection, consumer/contract law and product safety was covered above.

➔ Product safety

As explained above, **product safety legislation** requires producers to monitor all risks related to their products.

➔ Data Protection

There is no strict time limit on the obligation to take appropriate technical and organisational security measures (Art. 32 GDPR). This obligation **continues to apply** during all processing. The principles of processing in Art. 5 GDPR (especially purpose limitation and integrity and confidentiality) require that there is a process to **assess the security**. It can also be stated that the risks for which a DPIA has been carried out must be assessed continually (Art. 35(11) GDPR).

➔ Algorithmic trading

The obligation to provide resilient and adequately tested systems and the obligation to ensure business continuity arrangements have already been discussed.

Under Art. 14 of Delegated Regulation 2017/589, the business continuity arrangements shall effectively deal with disruptive incidents and, where appropriate, ensure a timely resumption of the algorithmic trading. These arrangements include a range of possible adverse scenarios in relation to the operation of algorithmic trading systems. The IT strategy discussed above also needs to be adapted to the business activities and risks to which the company is exposed.

	Did you inform end-users and subjects of existing or potential risks?

➔ General

These requirements should be read in conjunction with the obligations under Ethical Requirement 4: Transparency.

➔ Contract law

A seller is obliged to **indemnify the buyer against the hidden defects** of the sold item (Art. 1641 of the Civil Code). Under Belgian law, the seller is not bound by strict information obligations. Nonetheless, the seller must fully inform the buyer of what he is offering in good faith. The needs, expectations and possibilities of the buyer must be taken into account in this regard. Clear contractual agreements are therefore advisable.

➔ Safety at work

The Act regarding well-being at work and the Codex well-being at work also provide for employer obligations to ensure the safety of staff. This implies that the employer has to give all information concerning the risks and the preventive measures to reduce those risks (see also Art. I.2-16 Codex well-being at work).

➔ Algorithmic trading

The legislation on algorithmic trading requires investment firms to train staff on managing business continuity arrangements (see, inter alia, Art. 14 of delegated Regulation 2017/589).

	Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences? Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system?

➔ Cybersecurity

Various provisions on taking **technical and organisational measures** under the NIS Act have already been covered. The relevant actors should therefore identify all possible threats.

The Cybersecurity Act provides for additional rules regarding **certification**. In order to obtain a cybersecurity certificate, the entity developing the software is obliged to follow the **prescribed conformity assessment procedure**. Art. 51 provides several security objectives of cybersecurity certification schemes including detecting and documenting known dependencies and vulnerabilities and verifying that ICT products, ICT services and ICT processes do not contain known vulnerabilities.

➔ Product safety

The **general safety obligation**, the **monitoring obligation** and other obligations in specific regulations have already been discussed.

➔ Product liability

According to the Product Liability Act, a producer is liable for damage caused by **defective products** unless it proves, inter alia, that the state of scientific and technical knowledge at the time when the product was put into circulation did not allow the discovery of the defect's existence. A product is defective if it **does not provide the safety that can be reasonably expected**. A producer of products therefore has everything to gain from identifying the risks and threats in advance.

➔ Consumer protection

It has already been mentioned that the seller must inform the consumer in advance about the **functionalities of the product** or service, both for offline and online contracts. This implies that they have identified the potential threats.

➔ Contract law

The obligation for the **conform delivery** of goods and services as well as the information obligations of the seller and the service provider have already been discussed.

➔ Algorithmic trading

The obligation to provide sufficiently resilient and adequately tested systems has already been discussed. These include an IT strategy that is suited to the risks confronted by the business, meaning that the business must have defined those risks. As an arrangement for business continuity, the company also needs to describe the adverse scenarios associated with its operations among other things.

➔ Safety at work

The Act regarding well-being at work requires employers to take the necessary measures to promote the well-being of employees in the performance of their work. In this respect, the prevention principles to be applied are that the employer must prevent risks, assess risks that cannot be prevented and reduce risks as much as possible, taking into account technological developments. The Codex regarding well-being at work also provides for obligations with respect to the risk analysis (see inter alia Art. I.2.-5-I.2-7).

➔ Environmental safety

This must be read in conjunction with Ethical Requirement No. 6 Environmental and Societal Well-being.

	Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system?

➔ Cybersecurity

In the context of their overall obligations to provide for the security of their networks, both essential service providers and digital service providers are required to define **levels of severity** under the NIS Act. The Cybersecurity Act distinguishes between three levels of security levels: 'basic', 'substantial' or 'high'.

➔ Product safety

The risk-based approach to the **safety obligation** implies that certain products are subject to stricter rules than others. For certain categories (e.g. medical devices or certain types of machinery), stricter conformity assessment procedures may be imposed. This by itself implies that a distinction is made between the risks of different products. A **risk assessment** must also always be made when producers take corrective action, including recalls.

➔ Data Protection

The controller and the processor are required to implement appropriate **technical and organisational measures** to ensure a level of security appropriate to the risk. In assessing the appropriate level of security, account shall be taken in particular of the risks that are presented by processing (Art. 32 of the GDPR). A controller must also

notify a **personal data breach** to the supervisory authority (Art. 33) and to the data subject when it is likely to present a high risk to their rights and freedoms (Art. 34).

➔ Environmental risks

A detailed explanation of **environmental impact assessment, environmental impact reporting and product standards** can be found in the section regarding Ethical Requirement 6: Societal and Environmental Well-being.

	Did you assess the dependency of a critical AI system's decisions on its stable and reliable behaviour? Did you align the reliability/testing to the appropriate levels of stability and reliability?

➔ General

In general, AI systems are software applications with the ability to learn. Their **reliability** is therefore **essential** in all circumstances. Reliability means that the AI system does what it is designed to do. The reliability of software is more difficult to predict in practice for users. Indeed, software is not visible: the user only experiences the graphical interface but does not see any errors in the design of the software. This is even more difficult for AI systems, precisely because these systems autonomously alter themselves.

➔ Cybersecurity legislation

The aim of a **cyber security certification scheme** is to identify dependencies and vulnerabilities, among other things. If these are known, they must be documented. Furthermore, a 'Trustworthiness' standard is being developed by the ISO on this point.

➔ Product safety and product liability

The safety of products and services requires that, under normal or reasonably foreseeable conditions of use**, the product or service does not present any risk or only the minimum risks compatible with its use**, considered to be acceptable and consistent with a high level of protection for the health and safety of persons.

The Product Liability Act defines a defective product as one that does not provide the **safety that can be legitimately expected**. To accurately assess this, the **reliability of the system** must also be adequately assessed. The margin of error must be correctly estimated as well. Furthermore, the specific product regulations also provide for requirements on the reliability of the products (when they use AI for example). An example can be found in Annex 1 of the Royal Decree of 12 August 2008 on the placing on the market of machinery. This also provides for 'reliability' as one of the essential requirements.

➔ Algorithmic trading

The obligation to ensure the resilience of trading systems in Art. 22 of the Act of 21 November 2017 and the Royal Decree of 19 October 2017 has already been discussed. Further reference can be made to the operational requirements in delegated Regulation 2017/589, inter alia on business continuity arrangements, real time monitoring and security.

	Did you plan fault tolerance via, e.g. a duplicated system or another parallel system (AI-based or 'conventional')? Did you develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?

➔ **General**

In the information security assessment, **duplicated systems or parallel systems can be used** on which the necessary testing can then be done. Among others, this is how the LINDDUN threat modelling method works.

➔ **Product safety**

Art. IX.8 §2 of the CEL provides for the obligation of producers to take the **appropriate measures** to stay updated on the risks of these products and services, and to take the appropriate actions to prevent these risks. Distributors also need to remain updated on the product risks and report information to producers.

➔ **Cybersecurity**

The **obligations in the NIS Act** are not limited in time either. They therefore in principle apply throughout the activity of the provider of essential services or the digital service provider. However, these rules do not apply to micro or small enterprises as determined under Art. 32 NIS Act.

The Cybersecurity Act states that a cybersecurity certification scheme ensures that a range of security objectives are met. One of these is that ICT products, ICT services and ICT processes are provided **with up-to-date software and hardware that do not contain publicly known vulnerabilities**, and **are provided with mechanisms for secure updates** (Art. 51).

➔ **Consumer protection**

The relevant provisions regarding **security updates** under consumer regulations have already been discussed.

➔ **Personal data protection**

Various provisions on the **security of processing** in Art. 32(1) of the GDPR make it clear that the security obligation is permanent. This implies that the measures need to be evaluated at regular intervals for AI systems as well. A **DPIA** is also carried out not only at the start of any processing of personal data but also when there is a change of the risk represented by processing operations. This does not include an essential obligation to monitor regularly. Given the risks and unpredictability of AI systems, however, it is recommended that at least a system is in place to report new risks (Art. 35).

➔ **Algorithmic trading**

The general obligation for providers of algorithmic trading services to develop resilient systems has already been discussed. Regarding the technical requirements, reference can be made, inter alia, to the annual self-assessment obligations in Art. 2 of delegated Regulation 2017/589.

➔ **Environmental safety**

This is discussed in more detail under Ethical Requirement 6: Societal and Environmental Well-being.

3) Accuracy

	Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences? Did you put in place measures to ensure that the data (including training data) used to develop the AI system is up-to-date, of high quality, complete and representative of the environment the system will be deployed in?

➔ Data Protection

Under the **GDPR**, personal data **must be accurate and, where necessary, kept up-to-date** (Art. 5). At the very least, the personal data entered into the AI system must be **correct and accurate**. The controller is therefore in principle obliged to periodically update the personal data it processes. However, this is not always possible as information on accuracy will often come from the data subject, for example by communicating new data or making a request for rectification (Art. 16). However, any **rectification** must be communicated to each recipient of personal data (Art. 19). Furthermore, the process should be organised in such a way that the abovementioned **principles are respected** by virtue of the obligation to achieve data protection by design (Art. 25).

➔ Product safety

The safety obligation of producers (art. IX.2 CEL) means that the producer must ensure that the AI system does not contain **inaccuracies** that could lead to loss or damage. These risks must be permanently **monitored** (Art. IX.8 CEL). ISO also provides for relevant international standards for AI systems.

➔ Consumer protection

The requirement of **accuracy** may also form part of the objective conformity requirements within the Digital Content Directive (Art. 8) and the Consumer Sales Directive (Art. 7).

➔ Discrimination Law

For further explanation, see the discussion on Ethical Requirement 5: Diversity, Non-discrimination and Fairness.

	Did you put in place a series of steps to monitor, and document the AI system's accuracy?

➔ Data Protection

The **security obligation** in the GDPR has already been discussed. Among other things, this includes a process for regularly **testing, assessing and evaluating** the effectiveness of technical and organisational measures for ensuring the security of the processing (Art. 32 GDPR).

➔ Product safety

The **monitoring obligation** under Art. IX.8 §2 CEL has already been discussed. Producers must **monitor and document** the accuracy of the AI system. Taking samples and the monitoring obligation provide documentation that makes this possible. Distributors also have an obligation to monitor safety through various measures (Art. IX.8, §3).

Furthermore, both producers and distributors have an **obligation to report** to the Central Reporting Centre for Products (Art. IX.8, §4).

The specific product regulations also require the characteristics of each product to be identified as a minimum. Indeed, a common obligation is that the manufacturer must at least carry out a **conformity assessment procedure** or have one carried out. The manufacturer must also prepare **technical documentation**. This technical documentation must make it possible to assess the product's conformity with the relevant requirements (e.g. medical devices or machinery).

See in this regard also Art. 11 and 12 of the Proposal of a Regulation on AI on technical documentation and record-keeping requirements.

➔ **Cybersecurity**

The obligations of **actors to identify risks and take appropriate action** under the **NIS Act** have already been covered. Digital service providers are therefore obliged to keep track of and document all risks (Art. 30). They are also required to **report incidents** (Art. 35 et seq.). Actors must under the Cybersecurity Act also check any **dependencies** and **document vulnerabilities** (Art. 51). The certified manufacturer or provider of ICT products, ICT services and ICT processes must also **disclose** certain information (Art. 55).

➔ **Algorithmic trading**

The MiFID II legislation and the delegated Regulation also provide for a permanent self-assessment process, which requires providers of algorithmic trading systems to systematically assess whether their systems are sufficiently resilient. They are required to permanently adapt themselves in this regard.

	Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?

The **communication obligations** have already been discussed and are also covered in the discussion of Ethical Requirement 4: Transparency.

4) Reliability, Fall-back plans and Reproducibility

	Could the AI system cause critical, adversarial, or damaging consequences (e.g. pertaining to human safety) in case of low reliability and/or reproducibility?

The obligation to monitor reliability entails the same regulatory requirements as the **general security obligations and monitoring safety**. These have already been discussed.

	Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals?

➔ Data protection law

Personal data must be **adequate, relevant and limited** to what is necessary in relation to the purposes for which they are processed (Art. 5(1)(c) GDPR). Furthermore, the processing of personal data is only authorised if the processing is necessary for one of the legal grounds listed in Art. 6 GDPR. All these purposes must be listed in the **register** of data processing operations (Art. 32 GDPR).

➔ Product safety

In this context, the **monitoring obligation** discussed above is relevant. Noteworthy elements include the obligation to keep technical documentation and a register of complaints, as well as the general obligation to carry out spot checks to keep up-to-date on all the risks associated with products (Art. XI.8 CEL).

➔ Algorithmic trading

The obligations contained in the MiFiD II legislation have already been discussed above. As technical requirements, reference can be made to the obligations relating to annual self-assessment and validation, stress tests and the management of material changes.

	Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility?

The **general product safety requirements** apply in this regard. The **cybersecurity standards** are mainly focused on resisting attacks, not so much on flaws in the AI model. At first sight, no specific standards appear to be available for the reproducibility of AI systems.

	Did you put in place verification and validation methods and documentation (e.g. logging) to evaluate and ensure different aspects of the AI system's reliability and reproducibility? Did you clearly document and operationalise processes for the testing and verification of the reliability and reproducibility of the AI system?

➔ Data protection law

The GDPR requires a **record of processing activities** (Art. 30), a **notification of data breaches** (Art. 33), as well as the **prescribed measures** to be taken and documented. A **DPIA** can also be considered as a form of logging (Art. 35). See for more information Ethical Requirement 3: Privacy and data governance in this regard.

➔ Cybersecurity

The relevant provisions concerning the **identification of risks** and the taking of **appropriate technical and organisational measures** under the NIS Act have already been discussed (Art. 33). There is also an **obligation to report** incidents (Art. 35-37). The Implementing Regulation 2018/151 provides that the digital service provider may take a number of measures regarding the handling of incidents (Art. 2). The **certification scheme** under the Cybersecurity Act also tests for whether logging is in place to ensure that the security of the various systems is maintained (Art. 51).

➔ Environmental safety

More information can be found in the section on Ethical Requirement 6: Environmental and Societal Well-being.

➔ **Algorithmic trading**

In the context of the technical and organisational requirements for compliance with the security obligation under the MiFiD II legislation, reference can be made to inter alia the conformity tests as well as to the annual self-assessment.

	Did you define tested failsafe fall back plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?

➔ **General**

A failsafe is method that causes the AI system to **stop** or return to a safe state in the event of an error.

➔ **Standards**

At the IEEE level, a standard is currently being prepared for the failsafe design of autonomous systems: P7009 Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems. More information can be found here.

➔ **Algorithmic trading**

Section 3 of Delegated Regulation 2017/589 with regard to regulatory technical standards specifying the organisational requirements of investment firms engaged in algorithmic trading provides for various security obligations. For example, a **kill functionality** must be built in (Art. 12).

➔ **General safety legislation**

Producers are obliged to only place **safe** products on the market and offer safe services (Art. IX.2 CEL). In order to satisfy this safety presumption, harmonised standards that would apply to AI systems may be used (Art. IX.3 CEL).

➔ **Specific sectoral regulations**

Furthermore, some sectoral regulations may require a failsafe method to be envisaged as part of one of the **essential requirements**. Examples are included in the legal framework regarding machinery (point 1.2.1. of Annex 1) or medical devices.

➔ **Cybersecurity**

The NIS Act obliges the provider of essential services and digital service providers to take **appropriate measures** to prevent incidents or minimise their effects in order to ensure the continuity of these services (Art. 20 and 33). Implementing Regulation 2018/151 also contains provisions on **contingency plans and disaster recovery capabilities** (Art. 2(3)). These contingency plans make it clear that network and information systems must have a failsafe. ISO 27001:2017 is also relevant in this context.

The Cybersecurity Act also explicitly states as security objectives that in the event of a physical or technical incident, **access is restored** and ICT products, ICT services and ICT processes are secure by default and by design (Art. 51).

➔ **Data Protection**

The GDPR provides for an initial failsafe if there are errors in automated individual decision-making including profiling (Art. 22). Another failsafe is the exercise of the right to **rectification** (Art. 16) and to the **restriction of processing** (Art. 18).

	Did you put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score?

➔ Product safety

This obligation falls under the **general obligation** in accordance with Art. IX.8 CEL to take the necessary measures to remain up-to-date at all times with the risks applicable to the products and/or services that it puts on the market.

➔ Data protection law

Reference can be made to the **right to object to automated processing only, including profiling** (Art. 22 GDPR) and the **right to rectification** (Art. 16). Rules relating to the **DPIA** are also relevant (Art. 35 GDPR). These measures serve to prevent low reliability of the AI system or at least the consequences thereof.

➔ Consumer protection

Both the Digital Content Directive and the Consumer Sales Directive explicitly provide for the obligation to offer the **updates** that the consumer should expect.

➔ Cybersecurity

The definition of 'security of network and information systems' in Directive 2016/1148 points to **resilience**, but not necessarily to the reliability of the AI system as such. Other standards are therefore needed to complement this framework.

➔ Algorithmic trading

The MiFiD II legislation explicitly provides for the obligation to make effective arrangements in order to ensure, among other things, the continuity of services in the event of disruptions. If an AI system displays limited reliability, this is anticipated. In Delegated Regulation 2017/589, reference can be made to the control methods prescribed in Section II.

	Is your AI system using (online) continual learning? Did you consider potential negative consequences from the AI system learning novel or unusual methods to score well on its objective function?

➔ Product safety

Reference should be made to the **general safety obligations** in this regard (Art. IX.8 CEL). Throughout the lifecycle of the AI system, it must be ensured that the AI system is doing what it is supposed to do in the correct way. The question arises as to how this should be put into practice. A key element of this will be ensuring that the code of

AI systems is transparent at all times. See also the discussion under Ethical Requirement 4: Transparency. However, transparency, namely so-called 'explainable' AI, can also be a barrier to innovation. The right balance between transparency and functionality therefore needs to be struck.

Where are the possible points of improvement or focus points?

	**Convergence and coherence between different security regimes**

One of the main characteristics of (good) policy is **technology neutrality**. The rules must be applicable in any technological context and be resilient in the wake of technological change.

In the various legal areas that implement this ethical requirement, this is accomplished in a reasonably adequate manner. The safety standards are always linked to concepts such as 'appropriate' and 'safe', which in turn are linked to the risks known at the time of marketing and/or putting into use. These approaches – the European Commission's 'New Approach' and 'New Legislative Framework' – allow for flexibility: the general rules refer to standards drawn up by the market. This subsequently allows entities with relevant expertise to further develop the standards to deal with security risks. This is not necessarily linked to the product itself.

In relation to the continuous digitisation and the use of AI systems, **further convergence** between regimes is necessary. Multiple (safety) regimes apply to physical products and thus to AI systems embedded in hardware. The cybersecurity regime is primarily focused on preventing external attacks. However, it seems arbitrary to maintain a distinction based on the type of product and not on the risk. Furthermore, standardisation and certification for ICT systems in general, and for AI systems in particular, are still in their infancy (although much work is already being done within ISO and the IEEE). Only a few specific regimes such as the provisions on medical devices and algorithmic trading provide rules that do provide this convergence.

The **regulatory regime for safety must follow the current digital and AI-driven reality**. For the time being, this is only the case with certain types of applications, such as algorithmic trading. Some initiatives are already underway, such as the Digital Content Directive and the Consumer Sales Directive. Further development of the other regimes is necessary whereby the basis of the New Approach must be retained to ensure flexibility and legal certainty. **Interdisciplinary cooperation** between industry and academic institutions on the certification of AI systems is also useful.

There are few, if any, failsafe methods required or ways for consumers to obtain rectification. As with algorithmic trading, the priority is to develop **criteria** that determine when a failsafe or kill button is necessary. If the GDPR is applicable, a data subject can only exercise his/her right to object to automated processing including profiling. Given that this is limited to natural persons, companies cannot object if a decision that has legal consequences is made automatically. These consequences are already visible today in the automatic censorship of companies such as Facebook in their fight against fake news.

	**Access to technical standards**

Another area for improvement concerns **access to norms and standards**. Norms and standards are essential for developers to know what requirements their products or services need to meet. These standards are available for a fee through channels such as the NBN web shop. These costs can be prohibitive for some developers. It

would therefore appear appropriate to provide for greater access to technical standards and regulations, and to facilitate the possibility of consulting Belgian legislation. This is possible both through private initiatives (databases) and regulation.

	**Large responsibility on the part of the user**

There are various information, warning and safety obligations with respect to users. As such, there is a risk that users will be overwhelmed by the (amount of) information. This may eventually impede them to (properly) exercise their rights. Moreover, the end user is often a layperson so it is necessary to think about ways in which protecting the user can be organised (more) efficiently.

	**Clarification of legal concepts**

There is a need to **complement, adjust or clarify the applicable legislative framework**. The term 'product' does not explicitly refer to software in the applicable legislation. Irrespective of the question of the ultimate qualification of software, it is advisable to clarify this in the near future. A similar problem arises with respect to the concept of 'defect' and the 'legitimate safety expectations' of the general public with respect to AI systems.

What tools can be used to fulfil the ethical requirement?

➔ Cybersecurity

- Cybersecurity tool of the Centre for Cyber security Belgium to test how people comply with cybersecurity.
- The various threat modelling methods, including
  - Open Web Application Security Project provides a list of the 10 most frequent security threats to web applications;
  - LINDDUN.
- The frameworks relating to penetration testing
  - OSSTMM;
  - OWASP Web Application Penetration Checklist;
  - Penetration Testing Execution Standard.

➔ Standards

- Access via myNBN database to various standards applicable in Belgium.
- Access via the BBRI to standards related to construction products.

➔ Product safety

- The EU authorities provide a tool for the notification of dangerous products.

➔ Data Protection

- The CNIL provides a tool for conducting a DPIA.
- The Data Protection Authority provides a standard tool for reporting data breaches.

ETHICAL REQUIREMENT 3: PRIVACY AND DATA GOVERNANCE

What does the ethical requirement mean?

This ethical requirement relates to the fundamental right of every person to the protection of privacy and the closely related right to data protection. These rights help to protect the mental and physical integrity of natural persons. The principle of 'prevention of harm' therefore plays an important role in this ethical requirement. To limit this damage in practice, a **risk-based approach** is needed, which in itself requires high-performance data governance.

This ethical requirement covers the following **three sub-components** (1) privacy and data protection, (2) data quality and integrity and (3) data access.

Firstly, it concerns **privacy and data protection**.

AI systems generally require much information and data. This means that the collection, processing or generation of unlawful/unjustified information that belongs to the private sphere of the data subject or that provides an unwanted insight about the data subject needs to be avoided throughout the lifecycle of the AI system. The processing and generation of personal data must take into account the data subjects' right not to have this data processed.

Particular care must be taken not to draw conclusions from legitimately processed information that reveals aspects of a person's life which they do not wish to share or which could cause harm. Indeed, the processing of personal data can have a significant impact on a person's rights and freedoms and result in serious physical, material or immaterial damage (e.g. discrimination, damage to reputation, identity theft or fraud, financial losses or loss of confidentiality of personal data protected by professional secrecy). Compliance with the relevant rules is therefore essential to give citizens the necessary trust in AI systems. This will allow them to share their data with these systems in full confidence, knowing that they will not be unduly disadvantaged.

Secondly, this ethical requirement relates to the **quality and integrity of data**.

The quality of the data used to train AI systems is essential to ensure a qualitative outcome. In this regard, it is crucial that the data are accurate, sufficiently broad and representative, but also that they do not contain any socially constructed biases or inaccuracies. Data quality must also be ensured in the development and use phases by systematically testing and checking both the data sets and the processes. These checks should also be documented.

Finally, this ethical requirement relates to **access to data.**

The protection of personal data starts with a sound policy on access to personal data. By limiting access to the actual personal data to the individuals and situations in which such access is truly necessary, exposure will be limited and the risks of a deliberate and malicious data breach will be significantly reduced.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides questions based on the privacy/data governance section. Based on the same classification, we will for each question examine how they are (or are not) covered by existing regulations.

1) *Privacy*

	Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?

In view of the section under which this question falls, the right to physical, moral and mental integrity is interpreted as being related to or arising from (a violation of) the rights to privacy and data protection. Data protection is **highly regulated** and the obligation to lawfully process personal data is already included in a large number of national and international provisions.

➔ Fundamental rights

Art. 8 of the Charter of Fundamental Rights of the European Union (the 'Charter') and Art. 16 of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her.

Art. 7 of the Charter, Art. 17 of the International Covenant on Civil and Political Rights (ICCPR) and Art. 22 of the Belgian Constitution affirm the right to the protection of privacy.

Recognising the protection of privacy and personal data as fundamental rights already implies that the processing of information which may prejudice these rights requires a well-considered approach supported by law.

➔ European regulation

Data protection law has been largely harmonised at the European level by the **GDPR**. The GDPR governs the processing of personal data comprehensively and in detail. It is considered to be the 'basic law' for any processing of personal data. The GDPR follows a **risk-based approach**. This is expressed, among other things, in obligations to assess the risks prior to processing. It also has the general consequence that each processing operation must be assessed for its potential impact on the rights and freedoms of data subjects. The main purpose of these provisions is to protect data subjects (whose data are being processed) against violations of their rights and against harm.

The GDPR must be read together with the guidelines and positions of the European Data Protection Board (EDPB) and those of the supervisory authorities which can be considered as soft law. The decisions of supervisory authorities and national and European case law should also be taken into account in the interpretation and application.

According to the case, various provisions of the GDPR require developers, users and providers of AI systems to **think in advance** about the impact of their processing on the rights of a data subject. We refer in this respect to the previously published guide on data protection and AI, where much information can already be found. This guide covers a number of issues in detail. In the report at hand, we briefly discuss a number of important aspects.

Personal data can only be processed if the **principles of processing** such as 'lawfulness', 'transparency', 'purpose limitation' and 'data minimisation' are respected (Art. 5). These principles require that both the collection and the (further) processing of personal data are or will be carried out in accordance with these principles. The controller and processor must always be able to demonstrate that they are in compliance with their **obligations** under the GDPR ('duty of responsibility').

As such, the GDPR imposes as the core principle, inter alia, a **general transparency obligation** for any processing of personal data (Art. 5(1)(a), (b) and 12 and recital 58). This entails, among other things, a broad *prior* information obligation with regard to persons from whom personal data are collected and a similar information obligation with regard to persons whose data have been collected via third parties (Art. 12, 13 and 14 GDPR). For more information, we also refer to Ethical Requirement 4: Transparency.

Data subjects whose data are processed have, under certain conditions, a number of **specific rights** vis-à-vis the parties who collect and process their data even if such processing is part of the development, use or deployment of an AI system. Examples include the right of access, rectification, objection or not to be subject to a decision based solely on automated processing (Art. 15-22). These rights require the controller to ensure in advance that the personal data are processed in a manner that allows these rights to be fulfilled. The necessary processes will also have to be set up in advance in order to respond adequately to any exercise of rights.

The controller must implement **appropriate technical and organisational measures** to ensure and demonstrate processing in accordance with the GDPR (Art. 24(1) GDPR). In this regard, it must draw up an appropriate data protection policy if the scope of its activities so requires (Art. 24(2) GDPR). In our opinion, this would appear to apply at least to developers, providers and users acting as controllers. A controller must also implement appropriate technical and organisational measures both in determining the means of processing and in the processing itself (Art. 25). It also follows that processors should think in advance how AI systems may impact the various rights of data subjects and how to adequately safeguard them.

A **DPIA** includes the most concrete obligation to give prior consideration to the risks that the processing of personal data may pose to the rights and freedoms of natural persons. A DPIA must be performed in advance for any processing that is likely to present a high risk to data subjects (Art. 35(1) GDPR). This will generally always be the case when personal data is used in AI systems, at whatever stage. If there is any doubt as to whether performing a DPIA is necessary, a pre-DPIA can be performed beforehand. The 'DPIA manual' of the Belgian Data Protection Authority can be used for this purpose (Dutch only).

A DPIA describes the processing of personal data, assesses its necessity and proportionality, and helps to manage the associated risks to the rights and freedoms of natural persons by assessing these risks and defining measures to address them. As a result, organisations can check at the early stages of the development of an AI system whether there are discriminatory elements or whether certain individuals could be excluded. Although a DPIA is only mandatory for processing with a likely high risk, it is recommended to perform it in other situations as well. Indeed, it is a useful tool that helps organisations to comply with data protection legislation.

In addition, the controller may be required to consult the supervisory authority prior to processing if the processing involves a high degree of risk.

There is a mandatory **supervisory authority** in each Member State. For Belgium this is the Data Protection Authority or DPA (Art. 51 GDPR, see also the Belgian Act establishing the Data Protection Authority). The supervisory authority can sanction non-compliance with the GDPR, for instance by imposing a prohibition order and administrative fines. These fines can rise to €20 million or, if higher, an amount up to 4% of the total worldwide turnover in the previous financial year of the offender (Art. 84 GDPR).

This possibility of control and sanction requires that it can be demonstrated that the risks for the data subjects have been considered beforehand. This possibility will thus induce those who process personal data when using, offering or developing an AI system to actually document the risks in a substantiated way. The documentation can be submitted to the supervisory authority.

Furthermore, the GDPR contains a number of other relevant provisions that need to be considered *ex ante* to identify possible risks to the rights of data subjects (e.g. due to the processing of personal data by AI systems).

For example, the obligation to keep a **record of processing activities** (Art. 30), to **secure the processing** by appropriate technical and organisational security measures (Art. 32), to **notify data breaches** to the supervisory authority and possibly also to affected data subjects (Arts. 32 and 33), to appoint a **Data Protection Officer** or DPO (Art. 35), and to provide safeguards for transfers of **personal data to third countries** (Arts. 44-47)

➔ Belgian legislation

In addition to and in parallel with the GDPR, there are various **national provisions** that at least indirectly require parties to think about the impact an AI system can have on the relevant rights and freedoms of data subjects. The most important of these are briefly explained below.

The **Belgian Data Protection Act** 'implements' (among other things) the GDPR and lays down a number of rules that could be adapted or implemented by national law in accordance with the GDPR. In particular, the Data Protection Act provides for an exemption for public authorities as well as for processing operations outside the scope of the EU.

In addition, there are a number of **legal provisions** which govern specific aspects relating to certain processing of personal data. These are always aimed in part at protecting the rights of data subjects whose data is being processed.

These include the following legal provisions:

- The Belgian Act on the use of surveillance cameras;
- CLA no. 68 concerning the protection of the privacy of employees with regard to camera surveillance at work;
- The Patient Rights Act, which includes specific obligations for electronic communication service providers, but also general obligations that apply when cookies are used (e.g. in combination with an AI system aimed at direct marketing);
- the Electronic Communication Act;
- The Belgian NIS Act, which lays down a framework for the security of network and information systems of general interest for public security;
- provisions from the Code of Economic Law as they apply to consumer law (Book VI) or the law of the electronic economy (Book XII);
- the law on the national register, which prohibits, under penalty of criminal prosecution, the processing of a national register number without authorisation. This is therefore not possible in the context of AI systems either;
- CLA no. 81 protecting the privacy of employees with regard to the control of the electronic online communication data.

➔ Other standards

In addition to international and national rules, **other standards** may also encourage or require prior consideration of the risks that processing may pose to the rights of data subjects.

The **(non-binding) ISO standards** under the ISO 27000 series include information security standards. In addition, the ISO 27701 standard can be taken into consideration, which is an extension to ISO 27001 and ISO 27002. The application thereof also requires prior consideration of how certain processing operations may adversely affect the rights of data subjects. It also protects to a certain extent against the infringement of these rights.

The **PCI-DSS** (Payment Card Industry Data Security Standards) are self-regulatory information security standards imposed by the credit and payment card industry. These must be met by payment service providers in order to effect payments (e.g. in an online environment or in an application on a smartphone). They also require an in-

depth preliminary risk analysis. Any AI system that uses such a payment system, for instance in a user interface, or that may wish to use the data used in this regard must take this into account.

	Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?

From a regulatory perspective, the requirement to identify privacy issues during processing data is closely related to the requirement discussed above to **consider in advance the risks** that an AI system may pose to the rights of data subjects, notwithstanding the fact that in practice it is useful to ask both questions. For the applicable rules in this regard, reference is made to those discussed above. These also require that **privacy issues during the processing of data can be identified** and be **adequately followed up**.

2) Data governance

	Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?

This question is covered by the obligations arising from, inter alia, the **GDPR**, the Belgian **Data Protection Act** and **various other legal provisions** to comply with the obligations contained therein prior to any processing of personal data and during such processing. For a more detailed overview of the applicable standards, reference is made to the discussion above.

	Did you put in place any of the following measures, some of which are mandatory under the GDPR or a non-European equivalent? - Data protection Impact Assessment (DPIA); - Designate a Data Protection Officer (DPO) and include them at an early state in the development, procurement or use phase of the AI system; - Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications); - Measures to achieve privacy-by-design and default (e.g. encryption, pseudonymisation, aggregation and anonymisation); - Data minimisation, in particular personal data (including special categories of data).

Both the question of whether these measures should be taken and their possible implementation follow from the applicable data protection legislation such as the **GDPR**, the Belgian **Data Protection Act** and **various other laws** to perform a risk assessment when processing personal data and, if necessary, to take the measures corresponding to that level of risk. For a more detailed overview of the applicable standards, reference is made to the discussion above.

See in this regard also Art. 10 of the Proposal of a Regulation on AI dealing with data and data governance.

	Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system?

The obligation to implement these rights follows directly from the **GDPR** and is included in specific legislation for certain topics such as the **CEL** and the **Surveillance Camera Act**. For a more detailed overview of the applicable standards, reference is made to the discussion above.

	Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?

This question is covered by the obligations arising from, inter alia, the **GDPR**, the Belgian **Data Protection Act** and **various other legal provisions** to comply with the obligations contained therein prior to any processing of personal data and during such processing. This also applies to the personal data collected during the life cycle of the AI system. For a more detailed overview of the applicable standards, reference is made to the discussion above.

	Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data ?

This question is covered by the obligations arising from, inter alia, the **GDPR**, the Belgian **Data Protection Act** and **various other legal provisions** to comply with the obligations contained therein prior to any processing of personal data and during such processing. This requirement also applies to the personal data that can be created during the lifecycle of the AI system, although this is a question that will be easily overlooked in practice. For a more detailed overview of the applicable standards, reference is made to the discussion above.

	Did you align the AI system with relevant standards (e.g. ISO, IEEE) or widely adopted protocols for (daily) data management and governance?

This question relates to the voluntary application of **standards or protocols**.

There are no general mandatory standards or protocols applicable to the development, provision or use of AI systems.

For useful standards, reference can be made to the (non-binding) ISO standards under the **ISO 27000 series**, which address information security. In addition, the **ISO 27701** standard can be taken into consideration, which is an extension to ISO 27001 and ISO 27002. The **PCI-DSS** are also relevant.

Where are possible points of improvement or focus points?

Check-the-box attitude

A general problem in applying data protection law is the so-called check-the-box attitude, meaning that obligations are introduced in a **formalistic-administrative** way without actually contributing to adequate data protection or to a privacy-friendly (corporate) culture. The broader data protection policy is then (partly) reduced to a 'mere check', whereby all kinds of measures and processes are stipulated on paper but are not applied in practice.

Organisations should be encouraged to **effectively promote a privacy-friendly culture**. This can for example be achieved by organising workshops and setting up targeted information campaigns.

Lack of clarity

Despite the far-reaching harmonisation of data protection law through the GDPR, there are still many **questions about how certain principles should be applied**. As the GDPR is rather recent and technology continuously evolves, data protection law is still in full development. In September 2020 for example a draft directive was published on the concepts of controller and processor and another one on targeted advertising via social media.

At the same time, the interpretation of the GDPR is also influenced by (i) **various applicable national provisions**, (ii) sanctions, decisions and opinions of the national supervisory authorities, and (iii) evolving case law and legal doctrine.

In this regard, one should also consider that new provisions and guidelines based on the GDPR sometimes go against or impose additional conditions on best practices that already exist in these organisations. This becomes all the more complex in an AI context, given that many of these rules are difficult to apply in different AI contexts. Therefore, there is still a **need for clarification** on how to apply the GDPR in general and in an AI context in particular. As a step towards resolving this situation, we refer to our exploratory guide 'Artificial Intelligence and Data Protection'.

Insufficient knowledge in the start-up environment

In practice, start-ups are **often insufficiently aware** or **insufficiently preoccupied** with data protection. Workshops and targeted information campaigns could also be organised to remedy this shortcoming.

	**Lack of information and enforcement**

The Belgian Data Protection Authority has been actively carrying out inspections for over two years. Nearly 100 decisions have been issued in dispute cases (as of October 2020) and awareness-raising campaigns have been organised. However, it must be noted that the **likelihood of facing an actual sanction** in the event of a breach of data protection law remains very low. There is also limited awareness of privacy in the market, although it seems to be rising. The low likelihood of facing a sanction as well as the low privacy awareness undoubtedly contribute to the problems discussed above, which all too often means that data protection is not on the agenda at all or has low priority.

	**No obligation for developers or vendors to apply privacy-by-design to AI systems**

A key problem in applying data protection law in the technology sector, and therefore also in an AI context, is that it does in principle not apply to the developers and sellers of systems unless and to the extent that they act as controllers or processors. Indeed, only the two latter categories of actors have to comply with data protection rules and for example ensure that data protection by design and by default are applied.

Undoubtedly, it will be useful from a commercial perspective for a developer or provider to offer systems which comply with data protection requirements. However, the fact remains that the final responsibility lies with the user even though this is not always within their control. Ideally, the **obligation to ensure data protection by design and default would be imposed on all developers** of AI systems developed for or used within the European Union. This should ideally be done at the European level.

	**Slowing down the development and use of AI systems?**

According to some market players, the far-reaching obligations imposed by data protection law on the processing of personal data in an AI context slow down the development and efficiency of AI systems. As a consequence, society cannot fully exploit their potential.

It is therefore important, specifically for AI systems, to obtain **sufficient clarity** on how they should comply with data protection obligations, how these obligations can create an (unwanted) inhibiting effect and how they are compatible, for example with the 'black box' principle that applies to certain AI systems. As a step towards resolving this situation, we refer to our exploratory guide 'Artificial Intelligence and Data Protection'.

	**'Know your data'**

One question that is still missing in the assessment list but which is covered by data protection law is whether the processor or controller understands the **importance of the (personal) data used** and the **properties and attributes** that apply to them. This knowledge is required to ensure a good and appropriate quality of personal data. In itself, the 'know your data' principle constitutes the basic requirement for processing personal data in conformity with data protection law.

What tools can be used to fulfil the ethical requirement?

There are a large number of tools available to evaluate whether AI projects and systems comply with data protection requirements. Listed below are some publicly available tools published by public authorities, research institutes and other institutions.

- AI Blindspots to identify possible AI blind spots by reflecting on the decisions and actions taken prior to the development of an AI system.
- LINDDUN and LINDDUN GO are threat-modelling tools for privacy.
- Data protection impact assessments available online, such as this one from:
    - CNIL, the French supervisory authority: open source PIA software;
    - ICO, the supervisory authority of the UK: DPIA
- AI System Ethics Self-Assessment Tool allows users to assess how ethical an AI application is based on four ethical principles: fairness, accountability, transparency, explainability.
- Data Ethics Framework is intended to inform government agencies about the correct and responsible use of data when planning, evaluating and implementing policy or a service.
- Ethical OS tool kit is designed to help avoiding difficult-to-predict and unwelcome consequences when AI-based products and projects are developed.
- Data Ethics Canvas provides tools to address ethical questions during the design and development of a project or product based on data, such as an AI application.
- The Artificial Intelligence Impact Assessment
- The ECP's Supporting Ethics Approach (Dutch)
- Data Ethics Decision Aid (DEDA)

ETHICAL REQUIREMENT 4: TRANSPARENCY

What does the ethical requirement mean?

A crucial component of achieving trustworthy AI is transparency. Transparency encompasses different elements relevant to an AI system, including the **data**, the **system** and the **business models**.

This requirement of transparency consists of **three sub-components**, namely (1) traceability, (2) explainability, and (3) communication about the limitations of the AI system.

Traceability implies that the data sets and the processes from which the AI system's decision arises, including those of data collection and classification as well as the algorithms used, need to be documented as effectively as possible to make them traceable.

Explainability refers to the ability to explain both the technical processes of an AI system and the related human decisions (e.g. the scope of an AI system). Technical explainability requires that the decisions made by an AI system can be understood by humans and are traceable. When a decision by an AI system has significant impact on people's lives, it must always be possible to ask for a suitable explanation of the system's decision-making process.

Communication means that AI systems must not impersonate humans when interacting with users. People have the right to know that they are interacting with an AI system. This means that AI systems must be identifiable as such. Moreover, there must be an option for human interaction rather than interaction with an AI system when this is necessary to ensure compliance with fundamental rights. Furthermore, the capabilities and limitations of the AI system must be indicated with respect to both professionals and end users. The specific situation is therefore taken into account in this regard.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides several questions for each relevant sub-component. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration. Transparency requirements and information obligations play an important role in **different fields of the law** such as data protection law, consumer protection law, contract law and product safety rules.

1) Traceability

Measures must be put in place with regard to the traceability of the AI system throughout its lifecycle.

	Did you put in place measures to continuously assess the quality of the input data to the AI system?

➔ General

There are **various possibilities** that do ensure that the quality of input data and output for AI systems can be continuously assessed.

Contractual arrangements between the actors concerned would already seem to be appropriate in accordance with the applicable principles of contract law or specific legislation. The Act of 4 April 2019 on abuses of economic dependency, unfair terms and unfair market practices between companies is also important. This Act aims to improve the protection of (smaller) companies against larger players.

Other measures that can be implemented over time are also proposed by the AI HLEG, such as **standard automated quality assessment of data input**. This is possible by (1) quantifying missing values or gaps in the data, (2) detecting when data is insufficient for a particular task, or (3) verifying when the input data is erroneous, incorrect, inaccurate or not in the correct format.

Regarding the **quality of output data**, a number of measures/options are also proposed. These measures could for example take the form of a **standard automated quality assessment of the AI output** (e.g. the prediction scores are within the expected range, anomaly detection in the output and reassigning input data leading to the detected anomaly).

➔ Data Protection

The **GDPR** contains a number of relevant provisions. Personal data must be processed **lawfully, fairly and in a transparent manner**. They must also be accurate and, where necessary, **kept up to date** (Art. 5(1)). The data subject shall have a **right of access** (Art. 15) and the right to obtain from the controller without undue delay the **rectification** of inaccurate personal data concerning him or her (Art. 16). There is also a right to **restriction of processing** if for example the accuracy of personal data is contested (Art. 18). Under certain conditions, the data subject also has the right to **object**, on grounds relating to his or her specific situation, at any time to the processing of personal data concerning him or her (Art. 21) and **not to be subject** to automated processing, including profiling (Art. 22).

➔ Consumer protection

Directive 1999/44 (transposed in Belgium by the Act of 1 September 2004 and incorporated in Art. 1649bis-1649octies Civil Code), the Digital Content Directive and the Sale of Goods Directive may also be of interest because a defective quality of input data may result in a **lack of objective or subjective conformity of goods**, services or digital content, for which the seller can be held liable.

Producers are obliged to only place **safe products** on the market and offer **safe services** (Art. IX.2. CEL). If the safety of a product that incorporates AI is related to the input or output of data, this provision may be relevant. Sectoral legislation sometimes also imposes on producers the obligation for the product/service to be safe when placed on the market. This may be connected to the quality of their input data. For example, **medical devices** and **machinery**.

➔ Standards

In the future, information may also be found in **ISO standards** on AI. See in this regard also the discussion in Ethical Requirement 2: Technical robustness and safety.

	Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system?

➔ General

The contractual arrangements discussed above and the Act of 4 April 2019 are relevant.

➔ Data Protection

The GDPR also plays a role here. The controller has to take appropriate measures to provide the data subject with **all relevant information** regarding the processing in a concise, transparent, intelligible and easily accessible form, using clear and plain language. Among other things, this information pertains to the existence of automated decision-making, including profiling, and, at least in those cases, meaningful information about the logic involved as well as the significance and the envisaged consequences of such processing for the data subject (Art. 12-14). The data subject also has a **right of access** inter alia to the categories of personal data processed (Art. 15).

➔ Consumer protection

Consumer protection law also contains a number of rules regarding information obligations towards the consumer.

Pursuant to Directive 2011/83 on Consumer Rights, Art. VI.45 CEL stipulates **which information must be provided to the consumer clearly and understandably**. The information relates to the **main characteristics** of the goods or services, the functionality of digital content including applicable technical protection measures and the relevant interoperability thereof with hardware and software of which the seller is aware or can reasonably be expected to be aware.

In this context, reference can also be made to provisions in the CEL (Art. VI.93-100) on unfair and misleading commercial practices according to which information provided (e.g. on the operation of the AI system) must be **accurate** and **not misleading.** This provision is based on Directive 2005/29 on unfair commercial practices.

Directive 2019/2161 as regards the better enforcement and modernisation of Union consumer protection rules provides that consumers must be clearly informed when the **price they are shown is personalised** using **automated decision-making**. There is also an additional transparency requirement for online platforms regarding the display of search results. The consumer must be informed about the parameters used for the ranking of search results.

	Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system?

➔ General

The contractual arrangements discussed above and the Act of 4 April 2019 are relevant.

➔ Data Protection

According to the GDPR, the data subject has the **right to obtain confirmation** as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the listed information (Art. 15). The controller must also implement **appropriate technical and organisational measures** to ensure and be able to demonstrate that the processing is performed in accordance with the GDPR (Art. 24). The controller must also implement appropriate technical and organisational measures to comply with the

principles of data protection by design and by default (Art. 25). Each (representative of the) controller also needs to maintain a **record of processing activities** under its responsibility (Art. 30).

➔ Sectoral legislation

Sectoral legislation also contains provisions on drawing up, maintaining and updating technical documentation (e.g. the Medical Devices Regulation or the Machinery Directive).

➔ Liability

The implementation of adequate logging procedures will also become more important in the future within the adjusted legislative framework on liability. A '**logging obligation**' could in principle be imposed on producers of AI systems. This would imply that manufacturers equip their AI systems with the capacity to store all information that is usually essential to subsequently determine whether a risk associated with the technology has manifested itself (logging by design). Such a logging system would enable to identify the source of the malfunctioning that caused the damage. See in this regard also the logging obligations included in the Proposal of an EU Regulation on AI.

2) *Explainability*

We have already mentioned that explainability refers to the ability to explain the technical processes of an AI system and the reasoning behind the decisions or predictions that the AI system makes as well as the related human decisions. See in this regard also the transparency requirements imposed on certain AI systems by the Proposal of an EU Regulation on AI.

	Did you explain the decision(s) of the AI system to the users? Did you continuously survey the users if they understand the decision(s) of the AI system?

➔ General

This depends on the **way in which the explanation** can be given.

If the developers can directly interact with the users of the AI system, for example by organising workshops, this question/requirement can be explained there. However, if the developers are not directly involved with users, the organisation marketing AI systems must ensure that users understand the AI system and clarify any misunderstandings/uncertainties to developers.

➔ Data Protection

Under the GDPR, the controller has to take appropriate measures to provide **any information and any communication** relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language (Art. 12). In addition, controllers are also obliged to inform about the existence and use of automated (individual) decision-making and profiling, to provide useful information about the logic involved, and to inform about the significance and the envisaged consequences of such processing for the data subject (Art. 13(2)(f), 14(2)(g), 15(1)(h)).

➔ Product safety

Within the limits of their respective activities, producers must **provide users with the information** which will enable them to assess the **risks inherent** in a product throughout the normal or reasonably foreseeable period

of its use where such risks are not immediately obvious without adequate warnings, and to take precautions against those risks (Art. IX.8 CEL). This provision is useful for products that incorporate AI, such as autonomous vehicles.

Sectoral legislation may also be relevant. In accordance with Art. 5 of the Machinery Directive for example the manufacturer must provide the necessary information (e.g. the instructions) before placing a machine on the market and/or putting it into service.

3) Communication

The capabilities and limitations of the AI system must have been communicated to users in a manner appropriate to the use case at hand. This could include communication of the AI system's level of accuracy as well as its limitations. See in this regard also the transparency requirements imposed on certain AI systems by the Proposal of an EU Regulation on AI.

	In cases of interactive AI systems (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?

➔ E-commerce

Directive 2000/31 lays down standard rules in the EU for various matters relating to e-commerce. It also contains a number of **information obligations**. These must be provided in a **clear, understandable and unambiguous** manner before an order is placed. Prior to offering this service, a company must indicate, among other things, the various technical steps required to conclude the contract (Art. XII.7 CEL). As such, the fact that an AI system is used for concluding a contract must be indicated by a company allowing consumers to assess the possible risks of such systems.

For more information, see Ethical Requirement 1: Human agency and oversight.

➔ Modernisation of consumer law

Directive 2019/2161 regarding the better enforcement and modernisation of Union consumer protection rules also includes a number of provisions that are relevant. For example, consumers must be clearly informed when the **price** they see is **personalised** via **automated decision-making**. This will allow them to take potential risks into account in their purchasing decision.

	Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system? Did you communicate the benefits of the AI system to users? Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/ or error rates? Did you provide appropriate training material and disclaimers to users on how to adequately use the AI system?

➔ Consumer protection

The Consumer Rights Directive 2011/83 provides for enhanced consumer protection by harmonising key aspects of national legislation concerning contracts between customers and sellers. The Directive contains information

obligations. The Directive stipulates which **information must be provided to the consumer in a clear and understandable way**. This information relates to the main characteristics of the goods or services, the functionality of digital content including applicable technical protection measures and the relevant interoperability thereof with hardware and software of which the seller is aware or can reasonably be expected to be aware (Art. VI. 2, 1° and 8°, Art. VI. 45, 1° and 18° and Art. VI. 64, 1° and 17° CEL). Moreover, the information obligations concerning the functionality of the digital content must also allow consumers to consider the risks associated with the use of AI systems.

The **Digital Content Directive** and the **Consumer Sales Directive** are also relevant in this regard. Indeed, information obligations may be taken into account when determining whether a digital service or content or a product containing digital elements meets the contractually applicable subjective or objective requirements for conformity (e.g. functionality, compatibility and interoperability). In addition, goods and services with digital content must also be supplied with (installation) **instructions**. Finally, there are **update obligations** which aim at ensuring the continued conformity of digital content or services or of products with digital elements.

For more information, see Ethical Requirement 1: Human agency and oversight.

➔ Data Protection

According to the GDPR, any communication with a data subject must be in a **concise, transparent, intelligible and easily accessible** form, using clear and plain language (Art. 12). The purposes of the processing as well as the legal grounds and the legitimate interests for the processing must be clear. In order to ensure a proper and transparent processing, the existence of automated decision-making including profiling may also be provided. Useful information about the logic involved, as well as the importance and expected consequences of that processing for the data subject, may also be provided by the controller (Art. 13-14).

➔ Product safety

Within the limits of their respective activities, producers must **provide users with the information** which will enable them to assess the risks inherent in a product throughout the normal or reasonably foreseeable period of its use where such risks are not immediately obvious without adequate warnings, and to take precautions against those risks (Art. IX.8 CEL).

➔ Sectoral legislation

In accordance with the **Machinery Directive**, the manufacturer must provide the necessary information (e.g. the instruction manual) before placing a machine on the market and/or putting it into service (Art. 5). Pursuant to the **Medical Devices Regulation**, manufacturers have to ensure that, when placing devices on the market or using them, they are designed and manufactured in accordance with the applicable requirements. This includes providing the necessary **information obligations** such as labelling and instructions for use (Art. 10).

Where are the possible points of improvement or focus points?

	**Information obligations specific to AI systems**

There is already a wide range of legislation that applies to products that use/incorporate AI. Nevertheless, a **clearer framework** can be created through soft law and/or recommendations specifically aimed at information obligations regarding AI systems. Consequently, the legal framework does not need to be changed but could be

used as a basis to adopt specific information obligations for AI systems on (1) the origin and processing of input data, (2) the operation of the AI system, (3) how the system was created and (4) what the benefits and risks are. In this regard, it is necessary to identify the specific end user and to adapt the necessary information and transparency obligations accordingly. The idea could be to design a kind of '**AI leaflet**' that can be finetuned depending on the end user or customer. It should be determined how such a leaflet should look like, what it should contain, etc. In general, attention should be given to **legal design** so that the information is also understandable for laypersons. See in this regard also the transparency requirements imposed on certain AI systems by the Proposal of an EU Regulation on AI.

	**Documentation envisaged for (transparent) AI systems**

It must always be ensured that the relevant actors keep the necessary **documentation** on the functioning, decision-making process, etc. of AI systems. Regardless of the existing regulations, all information ranging from the design, development and use of AI systems should be archived in some way. One option would be to further encourage organisations to identify and evaluate the effects of AI systems on society, as is the case with a DPIA. In such an **Artificial Intelligence Impact Assessment**, the risks of using AI techniques for the users and organisations can be analysed. One could also think of introducing a **checklist** for the relevant organisations, companies and actors operating in the AI sector to increase the transparency of AI systems (e.g. identifying actors and determining how they are responsible for transparency or providing explanations).

	**Knowledge sharing and (multidisciplinary) cooperation**

It is important to **continue informing the general public about the reality and functioning of AI**. Simply **raising awareness** about the fact that AI systems are not always transparent (cf. black box) can already help. Users need to (continue to) know how AI systems work, which data is being processed, etc. Through collaborating within the Flemish AI Plan and Mediawijs as well as by participating in other projects (e.g. Iedereen Datawijs), the KDS already plays a role in this regard and must continue to do so.

Creating additional awareness regarding transparency amongst actors involved in the design and development of AI systems is also relevant. This can inter alia be achieved through the publication of best practices regarding the known vulnerabilities of AI models and technical solutions to address them. Formal **interdisciplinary cooperation** also needs to continue and be ensured, as is already the case within the Flemish AI Plan. **Information exchange** is crucial, also on an international level. Transparency in designing machine learning models can also be encouraged by continuously emphasising the need for an explainability-by-design approach for AI systems with a potentially negative impact on the fundamental rights of users.

What tools can be used to fulfil the ethical requirement?

There are several tools that have already been developed and are being used by AI developers to enhance the transparency of AI systems:

- Google Explainable AI: is a set of tools and frameworks to develop understandable and inclusive machine learning models.
- Microsoft Interpret ML: is a toolkit to help understand models and facilitate responsible machine learning
- LIME (Local Interpretable Model-agnostic Explanations): involves explaining what machine learning classifiers (or models) do.
- ALIBI: is an open-source Python library focused on machine learning model inspection and interpretation.
- DeepLIFT: Deep Learning Important FeaTures: is a method for decomposing the output prediction of a neural network on a specific input by backpropagating the contributions of all neurons in the network to every feature of the input.
- What If Tool: is a tool that can be used to test performance in hypothetical situations, analyse the importance of different data functions and visualise model behaviour across multiple models and subsets of input data.
- SHapley Additive exPlanations: is a game-theoretic approach to explain the output of any machine learning mode.
- AI Explainability 360: offers several algorithms that can be used to make explainability and fairness part of AI systems.
- Artificial Intelligence Impact Assessment
- The supporting ethics approach (Dutch, *Aanpak Begeleidingsethiek*)
- SDoC for AI / AI service FactSheets: asserts that suppliers of AI applications must create a fact sheet for each product, demonstrating that the application is 'compliant'. They do not provide a template, but they do provide some questions in the appendix as a basis.
- People + AI guidebook: contains principles designed to help user experience (UX) professionals and product managers take a human-centred approach to AI. This guide helps them put the user at the centre when developing an AI application.
- AI systems Ethics Self-Assessment Tool: is a questionnaire that helps to self-assess four ethical principles: fairness, accountability, transparency, explainability.
- The Principles for Accountable Algorithms and Social Impact Statement for Algorithms: consists of a number of questions that help ensure compliance with various ethical principles. The tool also includes steps that can be followed to further explore these principles in the different development phases of an AI project.
- TreeInterpreterRandom-Forest-Explainability-Pipeline
- Activation Atlases
- Rulex Explainable AI

ETHICAL REQUIREMENT 5: DIVERSITY, NON-DISCRIMINATION AND FAIRNESS

What does the ethical requirement mean?

In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. All relevant stakeholders must be taken into account throughout the process. There is also a need to ensure equal access through inclusive design processes as well as equal treatment. This requirement is closely linked to the principle of justice.

This requirement takes into account **three sub-components**: (1) avoidance of unfair bias, (2) accessibility and universal design and (3) stakeholder participation.

First, this requirement aims to **avoid unfair bias**.

Data sets in AI systems may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. This could lead to unintended (in)direct prejudice and discrimination against certain groups or people. This could potentially exacerbate prejudice and marginalisation. Identifiable and discriminatory bias should be removed in the collection phase where possible. There may also be bias in the way AI systems are developed (e.g. programming algorithms). This bias can be countered by establishing oversight processes to analyse and address the purpose, limitations, requirements and decisions of the system in a clear and transparent manner.

Secondly, it is related to **accessibility and universal design**.

Particularly in business-to-consumer domains, AI systems should be user-centric and designed in a way allowing all people to use AI products or services, regardless of their age, gender, abilities or characteristics. Accessibility to this technology for persons with disabilities, which are present in all societal groups, is of particular importance. AI systems should not have a one-size-fits-all approach and should consider Universal Design principles addressing the widest possible range of users, following relevant accessibility standards.

Thirdly, **stakeholder participation** is crucial.

In order to develop Trustworthy AI, it is recommended to consult stakeholders who may directly or indirectly be affected by the AI system throughout its life cycle. It is beneficial to gather regular feedback even after the deployment and to set up mechanisms that ensure stakeholder participation in the longer term. This can for example be achieved by ensuring that workers are informed and consulted and able to participate throughout the whole process of implementing AI systems at organisations.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides several questions for each sub-component. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration.

1) Avoidance of unfair bias

	Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design? Did you consider diversity and representativeness of end-users and/or subjects in the data?

➔ General

There are already several examples of AI systems that discriminate. For example, chatbot TAY rapidly became racist, Google Photos accidentally qualified black friends on a picture as gorilla's and Amazon's algorithm discriminated against women during a job application. There are no specific provisions regarding discrimination by AI systems (for the time being). Nevertheless, there is already a lot of relevant anti-discrimination legislation that may be relied upon in an AI context, especially since it is often people or human prejudice that underlie bias in AI systems.

The applicable and relevant legislation on **anti-discrimination is quite extensive** and can be found at European, national and regional levels. This legislation can be either general or specific and could be relied upon to develop specific legislation to mitigate and prevent **algorithmic discrimination**.

➔ General provisions on anti-discrimination: European Convention on Human Rights

There is an extensive body of European legislation on discrimination in general. The **European Convention on Human Rights** (ECHR) is enforceable and regulates the human and civil rights of all residents of the States that are party to the Council of Europe.

Art. 14 provides for a prohibition of discrimination. The enjoyment of the rights and freedoms set forth in the ECHR are secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

Case law from the European Court of Human Rights shows that both direct and indirect discrimination are prohibited. In the case of direct discrimination, someone is explicitly treated unequally on the basis of a protected ground such as nationality or gender. Indirect discrimination is a practice which at first sight appears neutral, but which in reality leads to discrimination against persons on the basis of a protected ground such as ethnicity. In the latter case, it is irrelevant whether there was an intent to discriminate or not. Instead, attention is given to the consequences of the practice.

➔ General provisions on anti-discrimination: European Union

The reference to non-discrimination can be found in the basic treaties of the EU:

- Art. 20-26 of the Charter of Fundamental Rights of the European Union: equality before the law, non-discrimination, cultural, religious and linguistic diversity, equality between women and men, the rights of the child, the rights of the elderly and the integration of persons with disabilities;
- Art. 10 of the Treaty on the Functioning of the European Union: combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation;
- Art. 2, 3(3) and 9 of the Treaty on European Union: focus on equality, non-discrimination, justice, pluralism, the rule of law and respect for human rights including the rights of persons belonging to minorities.

The EU has issued four general directives on discrimination:

- Directive 2000/43 of 29 June 2000 provides for equal treatment between persons irrespective of racial or ethnic origin;
- Directive 2000/78 of 27 November 2000 establishes a general framework for equal treatment in employment and occupation. The criteria are religion or belief, disability, age and sexual orientation;
- Directive 2004/113 of 13 December 2004 and Directive 2006/54 of 5 July 2006 (recast) provide for equal opportunities and equal treatment of men and women in matters of employment and occupation. The criteria are gender, including pregnancy and maternity.

Both direct and indirect discrimination are also prohibited at the EU level. In addition, the EU has issued a number of **specific directives**:

- Directive 79/7 of 19 December 1978 on the progressive implementation of the principle of equal treatment for men and women in matters of social security. The Directive applies to the working population;
- Directive 2010/41 of 7 July 2010 on the application of the principle of equal treatment between men and women engaged in an activity in a self-employed capacity;
- Directive 2010/18 of 8 March 2010 implementing the revised Framework Agreement on parental leave.

In addition, the provisions **regarding the processing of personal data** must be taken into account, which warn of possible discriminatory effects such as:

- The GDPR;
- Directive (EU) 2016/680 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties;
- Directive 2016/681 of 27 April 2016 on the use of Passenger Name Record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime.

➔ General provisions on anti-discrimination: Belgian legislation

These 4 European Directives were transposed at the federal level by the following **3 Acts of 10 May 2007**. They constitute the legal basis for combating discrimination.

- Act of 30 July 1981 punishing certain acts inspired by racism or xenophobia amended by the Act of 10 May 2007. This act prohibits discrimination on the basis of nationality, so-called 'race', skin colour, descent or national or ethnic origin;
- Act of 10 May 2007 on combating certain forms of discrimination. This act prohibits discrimination on the grounds of religion or belief, disability, age, sexual orientation, marital status, property, political opinion, trade union membership, language, current or future health condition, physical or genetic characteristic, social origin;
- Act of 10 May 2007 on combating discrimination between men and women. This act prohibits any form of discrimination based on gender. Discrimination based on gender reassignment, gender identity and gender expression is also covered.

➔ General provisions on anti-discrimination: Flemish legislation

In Flanders, anti-discrimination policy is laid down in a **number of decrees**:

- The Decree of 8 May 2002 on proportional participation in the labour market. This decree lays down that all groups of the population have the same right to participate in the labour market, regardless of the specific characteristics of certain groups of the population;
- The Decree of 10 July 2008 on a framework for the Flemish equal opportunities and equal treatment policy. This decree is the transposition of the 4 above-mentioned European Directives;
- The Decree on private employment mediation of 10 December 2010. This decree contains a limited number of provisions on discrimination (including legislation on interim agencies).

➔ Specific provisions on anti-discrimination

Besides these general provisions, there are also several **specific provisions** concerning discrimination. These are for instance included in Art. III.2 CEL concerning the **freedom of establishment**. The authorisation scheme may not discriminate against the service provider in question. Examples include better treatment for Belgian service providers or imposing additional conditions on foreign service providers. It also stipulates that recipients of services may not be subject to discriminatory requirements based on nationality or place of residence (Art. III.80-81 CEL). Examples include the reception of television services from another Member State, mobile telephony agreements or the purchase of goods and services on the internet.

Another provision relates to **access to payment accounts and basic banking services**. Consumers may not be discriminated against in any way on the grounds of nationality or residence, or of alleged race, colour, descent or national or ethnic origin (Art. VII.56/1 CEL). This obligation of non-discrimination applies when the consumer applies for, obtains access to or holds a payment account, or when they wish to use basic banking services. With regard to basic banking services, 'discrimination of any kind' is prohibited. This includes discrimination based on the financial situation of the consumer (Art. VII.57 §2 CEL).

There are also anti-discrimination provisions concerning the **conclusion of a loan agreement**. A lender is required to consult the Central Database to assess the creditworthiness of personal guarantors or consumers. The conditions governing access to the Central Database or any other database used to assess the creditworthiness of the consumer or a personal guarantor, or to check whether that creditworthiness is maintained, must not be discriminatory (Art. VII.77.1 § 1 CEL).

In the field of **insurance**, the general legal framework will apply. Insurance companies may use segmentation. This is a technique used by the insurer to differentiate the premium and possibly also the cover on the basis of a number of specific characteristics of the risk to be insured. The aim is to achieve a better match between the expected value of the loss and the costs to be borne by the policyholders' association and the premium to be paid for the provided cover. The Insurance Act does, however, state that any segmentation in terms of acceptance, pricing and/or the extent of cover must be objectively justified by a legitimate aim. The means of achieving that aim must also be appropriate and necessary (Art. 44).

	Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?

➔ Data Protection

Under the GDPR, any organisation that processes personal data must assess whether there are any risks involved. If an organisation suspects that an AI system is likely to pose a high risk to the rights and freedoms of natural persons, it must conduct a DPIA (Art. 35).

Moreover, data subjects can take action themselves by invoking their right of access and their right to rectification. The data subject has the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and some listed information (Art. 15). This allows data subjects to verify that this information is accurate and complete. If this is not the case, they can request the rectification of their personal data (Art. 16).

➔ Product safety

A number of relevant provisions are already in place for products or services that make use of AI (embedded software). This was covered in detail in previous requirements.

Following Art. 5(1) of the Product Safety Directive, Art. IX.8 §2 CEL stipulates that the producers of products and services must take measures within the limits of their activities that are commensurate with the characteristics of the products and services they supply in order to 1) **be informed of the risks** of these products and services and 2) take the **appropriate actions** to prevent the risks of these products. This includes a withdrawal from the market, appropriate and effective warnings to users and recalls. Art. IX.8 §3 CEL stipulates that distributors must contribute to compliance with safety requirements. This includes, within the limits of their activities, participating in monitoring the safety of products placed on the market.

Pursuant to Art. IX.8 §4 CEL, producers and distributors shall immediately inform the **Central Reporting Centre for Products** if they know or should know, on the basis of the information available to them and as a matter of professional judgement, that a product or service presents risks to the user that are incompatible with the general safety obligation. This obligation also applies to producers and distributors of products subject to specific safety legislation. Specific obligations are also sometimes imposed by EU regulations (e.g. Art. 11 Construction Products Regulation).

Both producers and distributors have an obligation to participate in the **continuous monitoring of product safety** throughout the life cycle of products. They have to make information available to users allowing them to assess the risks inherent in a product throughout its life cycle (Art. IX.8 of the CEL).

Both before and after placing a product or service on the market, the producer is obliged to test all possible **threats that could affect the product** in order to comply with the general safety obligation, the monitoring obligation and other obligations as stated in specific regulations. These obligations have already been discussed above.

Art. 8 of the **Product Liability Act** stipulates that a producer is liable for damage caused by defective products, unless he/she proves that it was impossible to discover the existence of the defect at the time the product was put into circulation, taking into account scientific and technical knowledge. A producer therefore benefits from identifying the risks and threats in advance.

The risk-based approach to the safety obligation implies that certain products are subject to stricter rules than others. For certain **categories of products** (e.g. medical devices or certain types of machinery), stricter conformity assessment procedures may be imposed compared to others. This in itself implies that a distinction is made between the risks of different products.

A risk assessment must also always be made when producers take **corrective action**. In practice, a risk assessment is always made when a corrective action is planned. This is clarified as it is explicitly stated that 'appropriate' actions must be taken to avoid risks from the products. In practice, a risk assessment will be carried out both when the product is placed on the market and when a corrective measure is taken. This depends on the corrective action. The basic principles of this risk assessment have already been explained in the European Commission's Guide to Corrective Action.

	Did you establish mechanisms to ensure fairness in the AI system?

➔ **General**

The development, installation and use of AI systems must be fair. The AI HLEG recognises that there are many different interpretations of fairness. However, the group argues that fairness has both a **substantive** and a **procedural dimension**.

➔ **Substantive**

This implies a commitment to ensure the **equal and fair distribution** of both benefits and costs, and to ensure that individuals and groups are free from unfair bias, discrimination and stigma. If unfair bias can be avoided, AI systems could even enhance social justice. **Equal opportunities** in access to education, goods, services and technology must also be fostered. In addition, the use of AI systems should never have the effect of misleading the (end) users or limiting their freedom of choice. Furthermore, fairness implies that professionals in the area of AI must respect the **principle of proportionality** between means and ends. They should carefully consider how to balance competing interests and objectives.

➔ **Procedural**

This includes the ability to **challenge** and **effectively appeal** decisions made by AI systems and by the persons who operate them. For this to be possible, the entity responsible for the decision must be identifiable and the decision-making process must be explainable.

➔ **Data Protection**

Personal data must be processed **lawfully, fairly and in a transparent manner** in relation to the data subject (Art. 5 GDPR)

	Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?

Although there is nothing relevant in the legislation in this regard, there are a number of **examples and practices** that want to increase awareness of designers and developers, such as the *Machine Learning Crash Course*, Google's *Responsible AI Practices* and several informative blogs. Ethical aspects are also extensively covered in various AI courses.

2) Accessibility and universal design

	Did you ensure that the AI system corresponds to the variety of preferences and abilities in society?

The UN **Convention on the Rights of Persons with Disabilities** contains a provision on accessibility. To enable persons with disabilities to live independently and fully participate in all aspects of life, States Parties have to take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, to information and communications, including information and communications technologies and systems, and to other facilities and services open or provided to the public, both in urban and in rural areas (Art. 9). Furthermore, a number of interesting documents on data quality and/or representativeness in an AI context have been published.

	Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?

➔ Treaty provisions

Once again, the UN **Convention on the Rights of Persons with Disabilities** contains a relevant provision. States Parties have to take appropriate measures to promote access for persons with disabilities to new information and communications technologies and systems including the internet. Appropriate measures also have to be taken to promote the design, development, production and distribution of accessible information and communications technologies and systems at an early stage. This will make these technologies and systems accessible at minimum cost (Art. 9).

➔ Public procurement

Directive 2014/24 on **public procurement** stipulates that for all procurement which is intended for use by natural persons, whether general public or staff of the contracting authority, the technical specifications shall, except in duly justified cases, be drawn up so as to take into account accessibility criteria for persons with disabilities or design for all users. These technical specifications need to lay down the characteristics required of a works, service or supply (Art. 42).

The Public Procurement Act of 17 June 2016 also contains a number of relevant **provisions on technical specifications**. In the case of public works contracts, a technical specification is defined as all the technical prescriptions contained in the tender documents that describe the characteristics required of a material, product or supply, so that it fulfils the use for which it is intended by the contracting authority. A design that meets all requirements, including accessibility for disabled people and suitability for use, is part of this. This also applies to public supply contracts and public service contracts (Art. 2, 44°).

➔ Accessibility of the websites and mobile applications of government bodies

At the federal level, Directive 2016/2102 was transposed by the Act of 19 July 2018 on accessibility of the websites and mobile applications of government bodies. Government bodies have to take the necessary measures to make their websites and mobile applications more accessible by making them perceivable, operable, understandable and robust. At the Flemish level, this Directive was incorporated into the Flemish Government Decree (Section 4. Accessibility of websites and mobile applications).

Directive 2019/882 on the accessibility requirements for products and services will also be relevant in the future. The Directive lays down accessibility requirements for important products and services such as telephones, computers, consumer banking services, electronic communications services including telephone and internet services, access to audio-visual media services and electronic commerce.

The Directive also sets common accessibility requirements for the design of the user interface and of the functionality of products. It also includes more specific accessibility requirements for certain electronic equipment for consumer use. For consumer products covered by the Directive, packaging, assembly instructions and other product information must be accessible. The Directive must be transposed by 28 June 2022.

	Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?

➔ General

The W3C Web Accessibility Initiative (WAI) develops standards and supporting materials to **help organisations understand and implement accessibility**. In this regard, a distinction is made between the concepts of accessibility, usability and inclusion. The European Directive on the accessibility requirements for products and services is also important in this respect.

➔ Standards

There are a number of **standards**.

- CEN/CLC/JTC 12 - Design for All
- European Committee for Standardization - Design for All specifies requirements that enable an organisation to design, develop and provide products, goods and services so that they can be accessed, understood and used by the widest range of users, including persons with disabilities.
- ISO/IEC 40500:2012 contains a wide range of recommendations for making web content more accessible. Adhering to these guidelines will make the content accessible to a wider range of people with disabilities.
- ISO/IEC TR 29138-1:2018 identifies a collection of user accessibility needs that diverse users have of ICT systems to make these systems accessible to them.
- ISO/IEC 30071-1:2019 sets requirements for the development process rather than the resulting products and services.
- W3C Web Content Accessibility Guidelines (WCAG) 2.0 provides recommendations for making web content accessible in general, and specifically for people with disabilities.
- Authoring Tool Accessibility Guidelines (ATAG) provide guidelines for the design of web content authoring tools that are accessible to people with disabilities and to help actors create more accessible web content.
- User Agent Accessibility Guidelines (UAAG) explain how to make user agents accessible to people with disabilities.
- ETSI EN 301 549 is suitable for public procurement of ICT products and services in Europe.
- Principles of Universal Design include the design of products and environments to be used by all people, as far as possible, without modification or specialised design.

Where are the possible points of improvement or focus points?

	**Stakeholder participation**

Although stakeholder participation is emphasised, the AI HLEG does not provide examples of such possibilities, nor does it elaborate on the concept. **Participation** is a value that is often identified but rarely further operationalised. It would help if there were concrete recommendations on how to engage stakeholders. There is also a need to work on developing the digital skills of the general population. This will ensure that everyone at least has a basic understanding of AI.

Inspiration can already be found in the practice of so-called **participatory design** in which users and stakeholders are involved throughout the entire development process of a certain solution/policy/product/service. It can be explored how that practice can also be relied upon in AI projects. Digital inclusion may be an additional focus for certain applications.

	**Discrimination**

Indirect discrimination can remain hidden for both the organisation and the victim. Self-learning algorithms are often 'black boxes'. Most people for example lack the expertise to understand how such systems make certain decisions. Even experts who develop the system do not always know how it will behave in reality. Due to the lack of transparency in decisions made by self-learning algorithms, it is difficult for people to ascertain whether they are being discriminated against.[2]

The regulatory frameworks on anti-discrimination protect individuals from discrimination on the basis of certain protected characteristics such as gender and ethnicity. However, algorithms can generate new categories of individuals based on seemingly innocuous characteristics, for example the choice of web browser or post code, thereby also potentially resulting in discrimination. **Algorithmic decision-making** can reinforce social inequality. Algorithmic pricing has in some cases for instance resulted in poor people being charged higher prices. More attention could/should be paid to this in terms of policy. The ability of consumer law to mitigate algorithmic discrimination could be assessed. Subsequent adjustments can be made if necessary.

Educational and awareness initiatives can be set up for developers of AI systems. In this regard, the focus should not only be on the typical grounds for discrimination (e.g. ethnicity, gender), but also on possible new forms of discrimination (e.g. based on browser behaviour) and the possibility that apparently 'innocent' data (e.g. postcode, diploma) could nevertheless lead to a form of discrimination. Instruments such as a DPIA or an AIIA are important to assess potential risks from the early stages in the life cycle of an AI system. The AIIA and the DPIA both use a risk-based approach and partly rely on the same logic. Both instruments are complementary but not interchangeable. A DPIA only addresses the risks that processing of personal data may entail for the data subject. The AIIA is a broader instrument that focuses on all possible ethical and legal issues which can be

[2] See for more information on this topic: Frederik Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision-making", COE, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 & Philipp Hacker, "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973.

associated with the application of AI. Moreover, the AIIA does not only look at risks but also provides a framework for making ethical decisions for the use of AI.

Policy makers may wish to consider whether and to what extent such an assessment would be appropriate for organisations. They can also check whether **adequate mechanisms are in place that indicate problems with bias**, discrimination or the poor performance of the AI system. If this is not the case, it can be examined whether the development of such a mechanism is possible and if so in which way. One can for example think of a point of contact to report problems related to AI systems.

	**Inclusion**

The AI HLEG highlights the needs of those with special needs or disabilities or those at risk of exclusion. Moreover, it also raises questions about the representation of different societal groups and the possible consequences of such representation regarding accessibility and inclusiveness. However, no questions are asked regarding possible access problems arising from financial criteria or technological illiteracy, which can also lead to exclusion. **Digital literacy** is becoming more important for the general public. The emphasis is often on technical skills and insights. Children are implicitly referred to but it is clear that in terms of access and design, a specific approach is required. It is also important that people become **aware** of the wider impact of AI on society as well as of its impact on their own lives, for example through the risk of discrimination. Policy makers can again play a role with regard to raising awareness and disseminating information.

What tools can be used to fulfil the ethical requirement?

- Ethical OS Toolkit: helps avoid difficult-to-predict and undesired consequences when AI-based products and projects are developed.
- Data Ethics Canvas
- Supporting ethics approach (Dutch)
- Artificial Intelligence Impact Assessment
- Aequitas: is intended to analyse whether there is bias in the data and models used. The audit can be done via desktop or an online tool.
- Building an Algorithm Tool: provides ethical questions that can be asked throughout the AI process (design, development, testing, and implementation).
- AI System Ethics Self-Assessment Tool allows users to assess how ethical a given AI application is based on four ethical principles: fairness, accountability, transparency, explainability.
- Unbias Toolkit: aims to share the online experiences of young people with policy makers, regulators and the ICT industry.
- Data Ethics Framework
- SDoC for AI / AI service FactSheets
- Ethics framework from Machine Intelligence Garage: consists of seven principles, each with a set of questions that can lead to a better understanding of how to address ethics in design.
- Data Collection Bias Assessment: provides that from the beginning of data collection, certain choices are fixed so that any bias can be detected at an early stage.

- AI Blindspots Map Set
- Tarot Cards of Tech
- AI Explainability 360: offers several algorithms that can be used to make explainability and fairness part of AI systems.

ETHICAL REQUIREMENT 6: SOCIETAL AND ENVIRONMENTAL WELL-BEING

What does the ethical requirement mean?

This ethical requirement is the expression of the ethical principles of **prevention of harm and fairness**. This requirement asserts that the broader society, other sentient beings and the environment should be considered as **stakeholders** throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged. Research should be fostered in this field.

This ethical requirement has **three sub-components**, namely (1) environmental well-being, (2) impact on work and skills and (3) impact on society at large or democracy.

AI systems must first and foremost be **sustainable and environmentally friendly**. AI systems are a double-edged sword when it comes to the environment. On the one hand, AI systems consume a lot of energy, which places an additional strain on the environment. On the other hand, AI systems can also be used to measure and recognise patterns in energy consumption and thus achieve a more efficient use of energy. The process of development, installation and use as well as the entire supply chain must therefore be monitored with a view to sustainability. Measures that promote the environmental friendliness of the supply chain of AI systems must be encouraged.

AI systems must also take into account the **social consequences**. Indeed, the use of AI systems can have an impact on our social relationships and attachment. AI systems can enhance social skills but can equally contribute to their deterioration. Physical and mental well-being can be affected in this respect. As such, the effects of AI systems must be carefully considered during the design, installation and commissioning. The ALTAI focuses on the impact on work and skills.

Finally, AI systems must **safeguard society and democracy**. They must maintain and promote democratic processes and respect the plurality of people's values and life choices. They must not undermine democratic processes, human debate or democratic electoral systems. AI systems must also be programmed in such a way that they do not operate in ways that undermine the basic principles of the rule of law and applicable laws and regulations. Due process and equality before the law must be ensured.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI includes several questions for each sub-component evoked. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration.

1) Environmental well-being

	Are there potential negative impacts of the AI system on the environment? Which potential impacts can be identified?

➔ Environmental quality standards

There is much environmental legislation. The general provisions can be found in the **Flemish Decree of 5 April 1995 containing general provisions on environmental policy**. Pursuant to Art. 2.2.1. et seq., the Flemish Government lays down **environmental quality standards** to protect the environment, which determine the quality requirements to be met by environmental parts/elements. There are two types of environmental quality standards: basic environmental quality standards and special environmental quality standards.

These environmental quality standards are set out in the Decree of the Flemish Government of 1 June 1995 concerning general and **sectoral provisions on environmental health** (VLAREM II). VLAREM also provides for different risk classes for activities. Establishments and activities are classified into classes 1 (highest risk), 2 or 3 (lowest risk) depending on whether the risk is higher or lower. **Monitoring** is essential for the enforcement of environmental quality standards in order to determine the state of all parts of the environment and to follow this up over time, as well as to identify any exceedances and their causes.

➔ Environmental and safety reporting

Following the Espoo Convention on environmental impact assessment in a transboundary context as well as Directive 2011/92 on the assessment of the effects of certain public and private projects on the environment or Directive 2001/42 on the assessment of the effects of certain plans and programmes on the environment, various provisions in Flemish decrees provide for a general obligation to **report on environmental effects**. **Safety reporting** aimed at preventing accidents is also relevant.

Title IV of the DABM provides the legal basis for all forms of environmental impact assessments. This is the procedure whereby prior to an application for a building or environmental permit for a given project or prior to a given plan, the **environmental consequences are studied and assessed**. As such, harmful effects on the environment can be assessed and tackled at an early stage.

Environmental impact assessments are also an important component of licensing procedures, such as the environmental permit. For certain projects, the government must first determine whether the project in question requires an environmental impact assessment.

Safety reporting is the procedure which may or may not lead to the drafting and adoption of a spatial planning safety report or an environmental safety report on a planned project and, where appropriate, to its use as an aid in the decision-making process concerning that project.

➔ Safety reporting on the operation of establishments

Following the Directive on the control of major-accident hazards involving dangerous substances (SEVESO III), the Cooperation Agreement of 16 February 2016 on the control of major-accident hazards involving dangerous substances provides for various obligations to be met by the operator of certain establishments in order to prevent major accidents involving dangerous substances. The operator must take all measures to prevent major accidents and limit their consequences for human health and the environment. Operators must also be able to demonstrate compliance with all obligations (Art. 5).

➔ Soil remediation

Implementing Directive 2010/75 on industrial emissions, the Soil Decree also provides for additional rules regarding the possibility of establishments and installations causing soil pollution. This is further implemented in the Decree of 14 December 2007 of the Flemish Government laying down the Flemish regulation on soil remediation and soil protection (VLAREBO). The decree provides, among other things, for rules on the

identification and inventorisation of contaminated land, the obligation to clean up and liability for that obligation, as well as rules on the transfer of land and the closure of establishments.

➔ Energy (and electricity in particular)

The main energy benefit of AI systems is that they can be used to improve overall energy efficiency.

Various regulations impose reporting of **energy performance** for certain installations. Similarly, different energy performance requirements are imposed. Many provisions in the Flemish energy performance regulations are the direct result of these requirements imposed by the EU (e.g. Directive 2010/31) (Art. 11.1.1. et seq. of the Energy Decree). Furthermore, an **energy performance certificate** (EPC) must be applied for buildings (Art. 9.2.1. Energy Decision). In the future, the use of AI systems in energy supply can be used to have a positive effect on this certificate.

It is crucial to accurately measure energy consumption to simultaneously identify the impact of AI systems on the environment and optimise energy consumption. This could involve AI systems in the form of **smart meters and smart grids**. Smart meters are digital meters that measure energy consumption in real time and transmit it to the grid operator. Smart grids are networks that use this information to bring electricity in real time to the places that need it most.

The use of smart grids and smart meters is widely encouraged in European regulations and policy documents (e.g. Art. 3.11 Electricity Directive). Furthermore, the Energy Efficiency Directive and Directive 2019/944 are also relevant. Flanders decided to introduce smart meters to measure electricity and natural gas consumption. The relevant provisions can be found in the Energy Decree and the Energy Decision. These documents also contain provisions on the processing of personal data that must be read in conjunction with the GDPR.

Large companies must undergo a **mandatory energy audit**. The relevant provisions can be found in Title 1 of the VLAREM (Art. 5 §8), Art. 7.7.2. of the Energy Decree and Chapter V of the Energy Decision. The rules provide for an obligation to draw up an energy plan, an energy study and an energy audit.

➔ Specific emission standards for vehicles

Given that **autonomous motor vehicles** are one of the more familiar AI systems, it is also appropriate to refer to the environmental requirements that apply to different types of vehicles. Think of Regulation 2019/631 of 17 April 2019 setting CO2 emission performance standards for new passenger cars and for new light commercial vehicles or Regulation 2019/1242 of 20 June 2019 setting CO2 emission performance standards for new heavy-duty vehicles.

➔ Product safety

The provisions on **product safety** in the CEL also overlap with environmental protection in some areas. Indeed, the risks to the health and safety of individuals also extend to protecting the environment, which in turn can harm the health and safety of individuals. An example can be found in the essential requirements for machinery as stipulated in Annex I of the Royal Decree of 12 August 2008 on the placing on the market of machinery. Risks due to the use of fuels, vibrations or radiation must be avoided when a machine is developed.

For more information, see also Ethical Requirement 2: Technical robustness and safety.

➔ Product standards

The **Rio Declaration** states that, to achieve sustainable development and a higher quality of life for all people, States should reduce and eliminate unsustainable patterns of production and consumption and promote appropriate demographic policies.

The Act of 21 December 1998 on product standards to promote sustainable production and consumption patterns and to protect the environment, public health and employees, provides for **specific standards on product quality**. All products placed on the market must be designed in such a way that their production, intended use and disposal do not endanger human health and do not contribute to an increase in the amount or harmfulness of waste and other forms of pollution, or contribute to a decrease thereof (Art. 4).

The King (i.e. the federal government) can take **additional measures** depending on this protection. Various additional royal decrees (RD) have been adopted. RD of 25 March 1999 lays down product standards for packaging. RD of 19 March 2004 lays down the product standards for vehicles. RD of 17 March 2013 provides standards limiting the use of certain hazardous substances in electrical and electronic equipment. Following the EU Packaging Directive, Art. 10 et seq. of the Product Standards Act provides for a prohibition to place products on the market with packaging that is non-reusable or not intended for recovery. In the implementation of the EU Directive establishing a framework for the setting of ecodesign requirements for energy-related products, Art. 14bis et seq. of the Product Standards Act provide for additional requirements for all products that consume energy.

➔ Eco-Management and Audit Scheme - the EMAS regulation and the mandatory environmental audit

Regulation 1221/2009 on the voluntary participation by organisations in a Community eco-**management and audit scheme** (EMAS) provides for a voluntary eco-management and audit scheme for all organisations. An organisation wishing to be registered under EMAS must carry out an initial environmental audit of all its activities, products and services. Pursuant to Art. 3.3.2. of the Decree on General Provisions of Environmental Policy, the Flemish Government may also designate the bodies obliged to perform a periodic or a one-off mandatory environmental audit.

➔ ISO 14000 family of standards

The ISO 14001-2015 standard lays down an international standard for **environmental impact analyses**. It is the first in a 'family' of standards that provides technical specifications for environmental management systems.

➔ EU Ecolabel

Furthermore, Regulation 66/2010 on the EU Ecolabel provides for a label for products: the **Ecolabel**. This system also aims to contribute to an increased level of protection via a label. This system has had limited success in practice to date.

	Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?

➔ Free access to environmental information

The EU Directive on public access to environmental information requires Member States to ensure that public authorities are required to make available **information relating to the environment** held by or for them to any applicant at their request and without having to prove an interest. Nonetheless, there are exceptions to this rule.

For Flanders, the transposition of the Directive can be found in Art. II.36 and further of the Government Decree, which provide for special provisions on access to environmental information. This information can be used in the context of AI projects to protect the environment. Bear in mind that the opposite is also possible: the protection

of trade secrets – an economic interest which is essential in all software activities – can also be used to block requests for environmental information.

➔ Environmental impact assessment reporting and environmental permit

The main provisions on environmental impact assessments have already been discussed in the previous question.

➔ Environmental management within companies

The Decree on General Provisions of Environmental Policy provides for various measures relating to **environmental management within companies** (Art. 3.1.1. et seq.). These imply that every company has to pursue sustainable production patterns and control. Companies also have to limit the environmental impact in all its aspects.

Some measures aim at reducing the environmental impact (e.g. appointing an environmental coordinator). Other obligations also include various forms of reporting that all companies – and therefore companies that develop, install or use AI systems – must comply with. The environmental audit has already been discussed in the context of the previous question.

➔ Soil remediation

This has also been discussed in previous questions.

➔ Waste management

Following among others the Waste Framework Directive, the **Materials Decree** provides for various obligations when waste materials are used. Natural and legal persons who manage waste are obliged to keep a chronological **register of wasteful substances,** which includes the following elements: the quantity supplied, nature, origin and, if applicable, the destination, frequency of collection, method of transport and treatment of the waste. Further rules are elaborated in the VLAREMA.

➔ Specific product standards for vehicles

Regulation 2019/631 requires an annual publication of the **environmental performance** of passenger car manufacturers. Regulation 2019/1242 also provides for specific **monitoring obligations** for manufacturers of heavy-duty vehicles.

➔ Energy law

The most important measures for individual users – the digital meter and the energy audit – have already been discussed. The energy performance certificate has also been discussed.

➔ Energy Report

Pursuant to Art. 12.1.1.-12.1.2. of the Energy Decree, the Flemish Government publishes an annual **energy report**. This contains comprehensive data on energy consumption for the Flemish Region.

➔ Product standards

This has already been discussed.

➔ ISO 14000 family of standards

This has already been discussed.

➔ EU Ecolabel

This has already been discussed.

	Did you define measures to reduce the environmental impact of the AI system throughout its lifecycle?

➔ Licensing obligation and the environmental permit

Since 1991, various projects and plans **are not permitted without a governmental license**. Pursuant to Art. 5.2.1. of the Decree on General Provisions of Environmental Policy, the Flemish Government draws up a classification list and determines whether the activity and establishment are subject to such a requirement (based on risks).

The previous system of government licensing was replaced in 2014 by the Decree on the environmental permit, along with Decision on the environmental permit. The **environmental permit** replaces the former environmental permit, the urban planning permit, the parcelling permit and the notification procedure. Applications are submitted to a single desk, the Environmental Desk, followed by a public inquiry and a consultation round. Without a prior environmental permit, it is not authorised to implement, operate, subdivide or make any change to a project which is subject to a permit requirement under the Flemish Code or the Decree on General Provisions of Environmental Policy.

➔ Environmental management within companies

The Decree on General Provisions of Environmental Policy also provides for additional obligations regarding environmental management within companies (e.g. appointment of an environmental coordinator). There is also an obligation to notify and alert the competent authorities in the event of accidental emissions (Art. 3.7.1.).

➔ Soil remediation

As already discussed, the Soil Decree provides for a **general remediation obligation**, which applies to the operator, the user and the owner of the land or establishment.

➔ Product standards

The legislation on product standards has already been discussed. The obligations under this Act ensure that producers must also **reduce** the potential risks of their products.

➔ Prevention of major accidents involving dangerous substances

As already discussed, the Cooperation Agreement on the prevention of major accidents obliges all operators to take the necessary measures to prevent major accidents and to limit their consequences for human health. In addition, the operator must also have identified and documented a **prevention policy** that ensures a high level of protection for human health. The operator must also incorporate an **internal emergency plan**.

➔ Waste management

Besides waste reporting obligations, the Materials Decree also provides for special provisions to make **waste processing more efficient**.

➔ Energy

Energy performance certificates for buildings, digital meters and energy audits have already been discussed. These measures not only serve to identify the impact of the establishments on energy consumption, but also provide an incentive to reduce consumption.

2) *Social consequences: impact on work and skills*

	In case the AI system interacts directly with humans:[3] - Did you assess whether the AI system encourages humans to develop attachment and empathy towards the system? - Did you ensure that the AI system clearly indicates that its social interaction is simulated and that it is not capable of 'understanding' and 'feeling'?

➔ Consumer protection and product safety

There are not yet specific forms of communication from AI systems that would indicate such behaviour. However, there is already a general framework for **transparency and information obligations** from consumer law and product safety law, among others. See in this regard also the transparency requirements imposed on certain AI systems by the Proposal of an EU Regulation on AI.

Given that AI systems can induce different forms of empathy, attention to **mental health risks** as well as further convergence between the hardware and software security regime is necessary to ensure an equivalent level of security in all cases.

➔ Personal data protection

Under the **GDPR**, there are also a number of provisions that are relevant in the context of the interaction between humans and AI systems. Many of these provisions have already been covered elsewhere. For example, the controller must process the data **lawfully, fairly and in a transparent manner in relation to the data subject** (Art. 5(1)(a)). There are also a number of information obligations relating to the existence of **automated individual decision-making** (Art. 13-14 and 22), as well as for the carrying out of a **DPIA** (Art. 35).

	Does the AI system impact human work and work arrangements?[4]

➔ General

Various potential risks posed by AI systems have already been identified. These systems could cause significant **disruption**, including to work and employment. The content of jobs and the different tasks involved may change due to increasing automation. However, AI systems can also improve safety in other areas. Robots can be used to make the lives of workers more pleasant or to keep them away from unsafe situations.

However, there are a **number of provisions** that must be taken into account when AI is introduced in the workplace. Human well-being is always the most important aspect in this regard.

[3] This question comes from the initial Assessment List, joined with the final draft of the Ethics Guidelines. The author(s) decided to do so as the initial version related to the impact on social skills more broadly, whereas the final version of the Assessment List relates strictly to the impact on work. The 'social impact' in both senses is relevant to evaluate. Insofar as the finalised version does not build upon the questions asked in the Assessment list joined in the Ethics Guidelines themselves, the latter Assessment List will be used to complement the finalised version.

[4] From here on out, we continue to assess the questions asked in the finalised Assessment List for Trustworthy AI.

➔ Human Rights

The international rights regarding work can be found in the **International Covenant on Economic, Social and Cultural Rights. An example is** the right to work and the right of everyone to the enjoyment of just and favourable work conditions (Art 6-9).

At the level of the Council of Europe, reference can be made to the **European Social Charter** which provides for the right to work (Art. 1), the right to fair working conditions (Art. 2), the right to safe and hygienic working conditions (Art. 3), the right to vocational training (Art. 10) and the right to health protection (Art. 11).

At European Union level, the **Charter of Fundamental Rights of the European Union** provides for the freedom to choose an occupation and the right to engage in work (Art. 15), as well as the freedom to conduct a business (Art. 16). It also provides for workers' right to information and consultation within the undertaking (Art. 27), the right of collective bargaining and action (Art. 28), the right of access to placement services (Art. 29), protection in the event of unjustified dismissal (Art. 30) and the right to fair and just working conditions (Art. 31).

The **Belgian Constitution** provides for the right to lead a life in accordance with human dignity. This right includes the right to work and free choice of employment and the right to social security (Art. 23).

➔ The Act regarding well-being at work and the Codex on well-being at work

There is also **specific legislation** on well-being and safety at work. The Act regarding well-being at work is the basic framework on health and safety at work. This Act creates a framework in which the implementing decisions are also taken. These implementing decisions are primarily brought together in the Codex on well-being at work. The codex is structured in an innovative way compared to the structure used in the General Regulations for Labour Protection (ARAB), the former codification of regulations on health and safety at work.

	Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?

➔ CLA No. 39

CLA No. 39 of 13 December 1983 concerning information and consultation on the social consequences of the introduction of new technologies provides for a specific arrangement requiring certain companies to **inform** employees if they wish to introduce a new technology. As such, it is useful to inform employees if there are plans to roll out an AI system in the workplace. The information to be provided includes, among others elements, the nature of the technology, the nature of the social impact and the time frame for implementation. The employer must also **consult** with the workers' representatives on the social consequences of introducing the new technology (Art. 2).

➔ Well-being legislation and Codex on well-being at work

Provisions on safety and well-being in the above-mentioned Act regarding well-being at work and the Codex on well-being at work should also be taken into account.

	Did you adopt measures to ensure that the impacts of the AI system on human work are well understood? Did you ensure that workers understand how the AI system operates, which capabilities it has and which it does not have?

➔ **CLA No. 39**

The **information obligations and the consultation** provided for in CLA No. 39 may be relevant in this regard (including on the nature of the new technology, on the health and safety of workers and on the skills and possible training and retraining measures for workers). These measures provide for an information obligation concerning the functionalities of the AI system and what capabilities it may or may not have.

➔ **Well-being at work**

Directive 89/391 on the introduction of measures to encourage improvements in the safety and health of workers at work lays down **general obligations** for employers to ensure the **health and safety** of workers (Art. 5). The Directive also provides for an **obligation to inform** workers about health and safety risks and protective and preventive measures and activities (Art. 10).

According to the Employment Contracts Act, the employer is obliged to ensure, with due diligence, that work is performed under **proper conditions** regarding the safety and health of the employee and that, in the event of an accident, first aid can be provided (Art. 20, 2°). As such, an employer must ensure that employees can interact with AI systems in a safe manner.

The Act regarding well-being at work also provides for specific obligations for employers to take the **necessary measures to promote the well-being of employees in the performance of their work** (Art. 5 §1). An employer must adapt the work to the individual in terms of the layout of workstations and the choice of work equipment and working/production methods to make monotonous and time-bound work more bearable while also limiting the consequences for health.

Furthermore, the Codex on well-being at work provides for additional obligations regarding **work equipment** (e.g. used machines, appliances, tools and installations). The employer must ensure that the work equipment is suitable for the work to be carried out so that the health and safety of workers is guaranteed during use (see, inter alia, Art. IV.2-1). The employer is also obliged to take the necessary measures to ensure that workers have at their disposal adequate information and, where appropriate, instructions concerning the work equipment used at work (see, inter alia, Art. IV.2-5). As such, the employer is already obliged to provide information on how to use AI systems under this legislation.

	Could the AI system create the risk of de-skilling of the workforce? Did you take measures to counteract de-skilling risks? Does the system promote or require new (digital) skills? Did you provide training opportunities and materials for re- and up-skilling?

➔ **CLA No. 39**

The **information obligation and the consultation** provided for in CLA No. 39 have already been discussed. If an AI system entails the risk that staff members no longer have the appropriate skills to do their job (de-skilling) the employer must inform the employees. The consultation must cover professional competence and possible measures for training and upskilling workers when AI is used.

However, there is no immediate question of actual reskilling under this CLA. The consultation obligation only applies in the event of 'significant collective social consequences'. In addition, the consultation obligation is not equivalent to an obligation to take effective action. CLA No. 39 does not require the employer to take effective measures to protect employees from de-skilling or to provide vocational training.

➔ **Obligations in terms of well-being at work**

Pursuant to Art. I.2-21 of the Codex on well-being at work, the employer is obliged to ensure that each employee receives **sufficient and appropriate training** in relation to the well-being of employees in the performance of their work, specifically tailored to their workstation or function. For instance, this training must be given when a new technology is introduced.

➔ **Vocational training**

The Flemish Service for Employment and Vocational Training (VDAB) was established by Decree of 7 May 2004. The specific organisation of vocational training is regulated by the Decision of 5 June 2009 on the organisation of employment services and vocational training. The VDAB has the power to propose to a non-working job seeker and compulsorily registered job seekers that they follow appropriate vocational training (Art. 66).

With respect to AI, the offer of training courses is limited at the moment. Yet, they do exist: the VDAB organises several training courses on artificial intelligence, both for job seekers and for employees.

3) *Impact on Society at large or Democracy*

	Could the AI system have a negative impact on society at large or democracy? Did you assess the societal impact of the AI system's use beyond the (end-)user and subject, such as potentially indirectly affected stakeholders or society at large?

➔ **General overview: human rights**

AI systems can have a significant impact on the **application of human rights** that aim to safeguard human dignity and our democratic system. The relevant human rights can be found in the International Covenant on Civil and Political Rights, the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union and the Belgian Constitution. Examples include the right to privacy, the right to freedom of thought, conscience and religion, the right to freedom of expression and the freedom of association. The use of social credit systems in China or the Cambridge Analytica scandal demonstrate that safeguarding fundamental rights is crucial and that people's autonomy can be affected. For more information on this subject, see Ethical Requirement 1: Human agency and oversight.

On the other hand, AI systems can also be used to disrupt the **functioning of democratic institutions**. AI systems are particularly well-suited to creating and disseminating fake or illegal information, which makes it more difficult to ascertain genuine information. Some examples include deepfakes or the ability for third parties to manipulate AI systems to spread false or harmful opinions (cf. racist chatbot TAY).

However, human rights are not absolute. The right to privacy, the right to freedom of religion and the right to freedom of expression in the ECHR may be subject to limitations insofar as these are prescribed by law, pursue a legitimate interest and are necessary in a democratic society to achieve that interest.

➔ Product safety

There are specific obligations for manufacturers to ensure the **health and safety** of all users when products are developed. The 'user' in this case is not just the consumer, but anyone who can use the product. Avoiding risks therefore implies avoiding risks to the wider environment of users.

➔ Standards

Recently, the IEEE also adopted a standard regarding impact of AI systems in general: IEEE P7010-2020 - IEEE Recommended Practice for Assessing the Human Impact of Autonomous and Intelligent Systems on Human Well-Being.

➔ Environmental law

The rules regarding the impact on the **environment and people's health** have already been discussed.

➔ Restrictions on surveillance and profiling

- Protection of personal data – restriction of processing of certain data

For any form of processing of personal data by private entities, the **GDPR** will of course apply. This was already discussed under Ethical Requirement 3: Privacy and data governance. Some provisions relevant to the creation of psychometric profiles and profiling in general are covered below.

Firstly, the principles of **purpose limitation, data minimisation and accuracy** must be observed at all times (Art. 5(1) GDPR). Furthermore, care must be taken with regard to the processing of **special categories of personal data**, including sensitive data and data relating to criminal convictions and offences. The processing of such information is only allowed in certain cases (Art. 9). Processing criminal data is only permitted under government supervision (Art. 10).

Art. 23 of the GDPR allows for certain **exceptions** (e.g. national and public security). These exceptions can also be found in the Act of 30 July 2018 on the protection of natural persons with regard to the processing of personal data.

- Personal data protection – profiling and automated decision making

Furthermore, **automated decision-making and profiling** are subject to special rules under the GDPR (Art. 13(2)(e); Art. 14(2)(g)). The data subject has the right not to be subject to automated processing only, including profiling (Art. 22). The data subject also has an explicit right to object to any processing for direct marketing purposes (Art. 21 of the GDPR). Furthermore, in certain cases it is necessary to carry out a **DPIA** (Art. 35(3) of the GDPR). Not coincidentally, these cases refer to situations where private surveillance is used and people are systematically and extensively profiled. Moreover, a **Data Protection Officer** must be appointed in certain cases. Again, it is no coincidence that situations were chosen where the controller or processor engages in surveillance or where the risk of surveillance is high.

- Protection of personal data – no transfer to third countries where different standards apply

Risks of human rights violations or disruptions to democracy can also come from countries outside the EU – countries where a different standard applies. The GDPR also provides special rules for the **transfer of data** to third countries (see e.g. Art. 44-46).

- Prohibition of content placed on the end-user's device

Following the e-Privacy Directive, Art. XII.12 of the CEL prohibits **storing text and other messages** on devices without the user's consent. An exception to this is direct marketing to existing customers. This exception can be found in the RD on the regulation of advertising by electronic mail.

- Third party processing by public authorities

Art. 10 of Directive 2016/680 provides that special categories of data may be processed if, subject to **appropriate safeguards for the rights and freedoms** of the data subject, and to the extent permitted by law, it is necessary to protect the vital interests of the data subject or if the processing relates to data which were manifestly made public by the data subject him or herself. This provision is repeated in the Data Protection Framework Act (Art. 34).

Profiling which results in discrimination against natural persons on the basis of their particular personal data is prohibited. Furthermore, any decision based solely on automated processing which produces an adverse legal effect for the data subject or significantly affects him or her shall be permitted only if a legal provision provides for appropriate safeguards with respect to the rights and freedoms, at least the right to human intervention (Art. 35). The transfer of personal data concerning a data subject for the purposes of investigation and law enforcement is permitted only under the conditions set out in Art. 66 and further.

- Restriction on the creation of certain AI systems (upcoming legislation)

Finally, the EU Commission's Proposal for an AI Act lists several new requirements to protect human beings from the most pernicious effects of AI systems. For example, AI systems that use subliminal techniques in order to manipulate persons, scoring AI systems or AI systems that misuse the vulnerabilities of specific groups of persons are prohibited.

➔ Criminal law restrictions on use of AI systems and dissemination of information through AI systems

- Prohibition on using AI systems to commit crime

AI systems **cannot be used as a means to commit crime** (see also Art. 66 and 67 Criminal Code). As such, it is illegal to develop AI systems that are intended to commit crime. Note, however, that this provision only applies to persons who develop AI systems intended to commit crime or people who train AI systems with the intent to commit crimes.

- It is also illegal to commit IT-related crime through AI systems

The criminalisation of the use of AI systems includes **IT-related crime**: forgery (Art. 210bis of the Criminal Code), IT fraud (Art. 504quater of the Criminal Code), as well as hacking and IT sabotage (Art. 550bis and 550ter of the Criminal Code).

- Criminal law prohibits certain forms of expression

The principle is that people are free to have and express their opinions. In this sense, people should be allowed to say whatever they want on the internet. It is also allowed to simply share things that are disseminated by an AI system. However, this freedom is not unlimited.

Criminal law **prohibits various forms of opinions** that can be shared. In our view, a number of these may be relevant to AI systems (e.g. fake news, deepfakes): IT forgery (Art. 210bis Criminal Code), slander, defamation and insult (Art. 443 et seq. Criminal Code), harassment (Art. 442bis Criminal Code), incitement to racism (Art. 20 and further Antiracism Law), dissemination of sexist opinions (Art. 2-3 of the Act of 22 May 2014 to combat

sexism in public spaces), pornography (Art. 383bis Criminal Code), sharing of image or sound recordings of an exposed person (Art. 371/1 Criminal Code).

➔ Self-regulation in the press

The freedom of press is a fundamental principle subject to the limitations mentioned above. Besides criminal law, the self-regulatory measures issued by the Council for Journalism also apply in principle. Written publications are subject to the **Code of the Council for Journalism**. This is not a binding regulation but a guideline that in principle applies to all publications, regardless of the medium. This code is a form of self-regulation and is therefore not directly enforceable in a court.

➔ Media law obligations for audiovisual messages

In implementing the **Audiovisual Media Services Directive**, the Media Decree contains various obligations for all broadcasters. It also contains provisions applicable to online platforms on which users upload videos, such as Twitter, YouTube, etc. Art. 37-38 of the Media Decree stipulate that freedom of expression is guaranteed and that broadcasting activities cannot incite people to hatred or violence. However, media activity has shifted to these platforms. This means that many 'filters', applied by editors for example, do not automatically apply to these platforms. These obligations will at some point be complemented by the obligations imposed by the Digital Services Act Proposal on online platforms and online service providers, once it has been adopted.

	Did you take measures to minimise potential societal harm of the AI system?

➔ Protection of personal data (risk of surveillance and manipulation)

This has already been discussed above.

➔ Safety and environment

These have already been discussed above and under Ethical Requirement 2: Technical robustness and safety.

➔ Enforcement

Reference can also be made to the discussion under Ethical Requirement 7: Accountability.

	Did you take measures that ensure that the AI system does not negatively impact democracy?

➔ Safeguarding elections

We can conclude that the risks of for example **voting fraud** by AI systems are limited in Belgium. The reason is that the current voting system is still almost entirely based on paper voting or, if voting is automated, on specific voting computers that are not connected to a network. The specific provisions can be found in the Decree of 25 May 2012 on the organisation of digital voting in local and provincial elections ("Digital Decree") and the Act of 7 February 2014 on the organisation of electronic voting with paper evidence.

➔ Risk of concentration and response from competition law

Another risk of the increased use of AI systems and digitisation in general is that only a **handful of tech companies** – Facebook, Google, Amazon, Apple and Microsoft – have the power to determine what we see and hear. According to some, these tech companies are becoming so large that they are gradually outgrowing the states that are supposed to regulate them. For this reason, there is a fear that they will be able to use their market power to steer the social debate as they see fit, partly through the use of their algorithms (cf. surveillance capitalism). This growth is also very difficult to stop. WhatsApp and Instagram for example are already owned by Facebook. That is why **competition law** is crucial. Competition law contains rules that ensure that free competition between companies is upheld. This also benefits the functioning of democracy.

Competition law is applied at two levels in Belgium. At the federal level, the provisions are found in Book IV of the CEL. Enforcement is the responsibility of the Belgian Competition Authority. Furthermore, competition law is regulated at the European level for all matters affecting trade between the EU Member States. These rules can be found in Art. 101 and further of the Treaty on the Functioning of the European Union and the transposing regulations. In principle, enforcement is the responsibility of the European Commission, with a possible review by the Court of Justice of the European Union. The Commission has already fined Google for linking its search engine and browser to the operating system of its smartphones and for abusive practices in managing its search engine.

➔ **Restrictions on surveillance**

These have already been discussed.

➔ **Restrictions within criminal law to safeguard the social order and democratic debate**

These have already been discussed.

➔ **Rules on the liability of online service providers**

Following Art. 12 to 15 of the E-Commerce Directive, Art. XII.17 to XII.20 of the CEL provide for a **general exclusion of liability for online intermediaries**. This refers to all IT services in which these intermediaries play a purely passive role (cf. mere conduit providers, caching providers and hosting providers).

As a result of the increasing dissemination of illegal information, harmful information and false information, these rules are coming under increasing attack. In its Communication 'Tackling illegal content online', the European Commission stated that online platforms should take **effective proactive measures** to detect and remove illegal online content and not only react to reports that they receive. This suggests that online platforms will be required to filter illegal content themselves in the future. More recently, these rules are undergoing reform. On 15 December 2020, the European Commission published its Digital Services Act Proposal, which resumes the liability exemptions, but also complements them with a permission to block certain types of content in good faith.

Where are the possible points of improvement or focus points?

Comprehensive regulation on the environment and energy

For the environment and energy, existing **legislation** already subjects all-encompassing risks to governmental oversight. These regulations also impose obligations regarding the design and use of products as well as obligations for establishments and installations. The rules contain rather extensive procedures based on general

standards. These standards are not necessarily specifically tailored to AI systems and it is also not required. The regulations should be based upon the risks that all installations and/or processes have on the environment. However, certain issues have been identified that require **further convergence** as a result of AI systems. For example, product standards legislation exclusively focuses on tangible movable goods. These legal frameworks would therefore only apply to AI systems if they are embedded in hardware.

	**Energy efficiency must not come at any cost**

Using AI systems to improve **energy efficiency** must ensure that all other rights of users are respected, including their rights to security and privacy. A general problem that arises when entrusting interconnected systems is that they become more susceptible to external manipulation.

	**Use of AI systems, social skills and labour**

Consumer protection law and the rules on the processing of personal data already impose information obligations regarding the **impact that AI systems can have on our social lives**. With regard to labour, there is a limited **legal framework** which sufficiently emphasises the need for training, consultation and safety. However, continuous investment in the necessary training courses/upskilling and acquiring digital skills is required. In other words, it has to be ensured that everyone is involved in the digitisation/AI train.

	**Law enforcement on the web**

The functioning (and hence consequences) of AI systems does (do) not stop at national borders. This can create problems as the jurisdiction of governments does stop at national borders. This makes it difficult to tackle perpetrators of criminal acts on the internet. There needs to be a continuous focus on preventive measures (e.g. reinforcing the filtering obligations of platforms and information sharing companies).

What tools can be used to fulfil the ethical requirement?

The **tools already covered** in the other ethical requirements are relevant to ensure reliable AI. As a result, the social impact is also taken into account. Moreover, there are a number of specific courses that help raise societal awareness on AI. Examples include:

- Elements of AI
- Flemish AI course
- AI in business (Agoria)

ETHICAL REQUIREMENT 7: ACCOUNTABILITY

What does the ethical requirement mean?

The requirement of accountability complements the above mentioned requirements and is closely linked to the principle of fairness. Under this requirement, mechanisms need to be put in place to ensure **responsibility and accountability for AI systems** and their outcomes, both before and after deployment. As such, the necessary measures need to be taken to ensure and encourage accountability when developing or using AI systems. This requirement implies that the potential risks of AI systems are identified and mitigated in a transparent manner. When unjust or adverse impacts occur, accessible mechanisms for accountability should be in place that ensure an adequate possibility of redress. In other words, it needs to be ensured that someone can be held responsible if AI systems cause harm and that adequate compensation is envisaged.

The ethical requirement consists of **two sub-components** (1) auditability and (2) risk management.

Auditability means that it is made possible to audit the algorithms, data and design processes. This does not necessarily imply that information about business models and intellectual property related to the AI system must always be openly available. The possibility for both internal and external actors to conduct evaluations as well as to access records on said evaluations can contribute to Trustworthy AI. In applications affecting fundamental rights, including safety-critical applications, AI systems should be able to be independently audited.

Risk management means that the ability to report on actions or decisions that contribute to the AI system's outcome and to respond to the consequences of such an outcome must be ensured. Identifying, assessing, documenting and minimising the potential negative impacts of AI systems is especially crucial for those (in)directly affected. Due protection must be available for whistle-blowers, NGOs, trade unions or other entities when reporting legitimate concerns about an AI system. When adverse impact occurs, accessible mechanisms should be foreseen that ensure adequate redress.

Which rules are an expression of the ethical requirement or can serve as inspiration?

The ALTAI provides several questions for each sub-component. We will give an overview of the questions per sub-component and any relevant legislation that already reflects these questions or that can be used as inspiration.

1) Auditability

	Did you establish the mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact)?

➔ **General**

This requirement should be read together with the discussion of Ethical Requirement 4: Transparency. **Contractual arrangements** between the actors involved are also appropriate to provide the necessary guarantees of auditability.

➔ Data Protection

There are various provisions in the **GDPR** regarding transparency (Art. 12 and 13-14) and the data subject's right of access (Art. 15) which have already been discussed elsewhere. The obligation of the controller to keep a record of processing activities, including a general description of the technical and organisational security measures if possible (Art. 30), was also discussed. Any organisation that processes personal data must assess whether there are any risks involved. If an organisation believes that an AI system is likely to pose a high risk to the rights and freedoms of natural persons, it must conduct a DPIA.

➔ Consumer protection

The **information obligations** in Directive 2011/83 on consumer rights have already been discussed and are also relevant in this context. It must be specified whether there is a possibility of having recourse to an out-of-court complaint and redress mechanism, to which the trader is subject, and the methods for having access to it (Art. 6 and Art. VI.45 CEL).

The implementation of **adequate logging** procedures will presumably gain importance in the future, for example within the forthcoming framework on liability and AI systems and the Proposal of Regulation on AI.

Within the **CEL**, a number of provisions can be found that are relevant to ensure/simplify the auditability of the AI system. For example, within the limits of their respective activities, producers need to provide users with the information which will enable them to **assess the risks inherent in a product** throughout the normal or reasonably foreseeable period of its use, where such risks are not immediately obvious without adequate warnings, and to take precautions against those risks. Within the limits of their respective activities, producers take measures commensurate with the characteristics of the products and services that they supply in order to 1) be informed of risks posed by such products and services, and 2) take appropriate action to prevent those risks, including withdrawal from the market, adequately and effectively warning users and recall from consumers (Art. IX. 8).

Producers and distributors must also immediately inform the Centraal Meldpunt voor Producten if they know or should know, on the basis of the information available to them and as a matter of professional judgement, that a product or service they have placed on the market presents risks to the user that are incompatible with the general safety obligation (Art. IX.8 §4 CEL). They have to provide specific information (e.g. data allowing exact identification of the product and all available information allowing the product to be traced).

In addition, many **sectoral regulations** in which measures are imposed that ensure/simplify the auditability of products using AI (e.g. technical documentation) are also relevant. For example, the Medical Devices Regulation (Art. 10(4), 10(5), 10(9) and 10(13)). The regulations on machinery also state that the manufacturer must ensure that the technical file is available before placing a machine on the market and/or putting it into service (Art. 7). Also, under the NIS Act, technical and organisational security measures must be taken to prevent incidents or limit their impact (Art. 20-23).

	Did you ensure that the AI system can be audited by independent third parties?

The **certification** of AI systems has already been emphasised in various policy documents. For example, there are calls to make high-risk AI systems subject to mandatory certification. Voluntary certification has also been put forward as an option for regulating AI. The Proposal for a Regulation on AI also contains provisions on third-party certification and conformity assessment.

However, there are many other ways to obtain labels.

- The **CE mark** for example indicates that according to the manufacturer a product meets all EU requirements in terms of safety, health and environmental protection. The marking is mandatory for certain categories of products sold in the EU, even if they are manufactured elsewhere. CE marking is only mandatory for products for which EU specifications exist and, in addition, those specifications must require that the products display the CE marking. For certain products, an independent notified body has to be involved in the assessment of whether a product is in conformity and can therefore be given the CE mark. This can be found in rules per product category. If a product does not have to be assessed by an independent body, it is up to the producer to check whether it meets all technical requirements. This means that the possible risks of use of the product are assessed and documented.
- **Sectoral provisions** play an important role as well in determining whether certification or another conformity assessment is necessary (see e.g. Art. 52-60 Medical Devices Regulation).

2) *Risk management*

	Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures? Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system? Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas? Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?

➔ General

The aim is to **identify** the ethical aspects of AI in design, development and use. In this way, reliable AI systems can be developed.

There is currently not much legislation on this subject, precisely because the concept of ethics and its 'legal translation' is not always easy. However, applicable regulations do sometimes require actors to conduct a risk analysis that can be used as a starting point and source of inspiration in an AI context. One example is the obligation under the **GDPR** to perform a DPIA if an organisation considers that an AI system is likely to pose a high risk to the rights and freedoms of natural persons.

➔ Sectoral legislation

Inspiration can also be found in **sectoral legislation**, for example in the Medical Devices Regulation through the quality management system (Art. 10). The system includes all parts and components of the manufacturer's organisation which deal with the quality of processes, procedures and devices. It lays down the structure, responsibilities, procedures, processes and management resources required to implement the principles and measures necessary for compliance with the provisions of the Regulation.

The product safety regime in the CEL also contains a number of relevant provisions with regard to taking the necessary measures to be able to stay informed about the risks of products and services and to take the appropriate actions to prevent these risks.

	Did you establish a process for third parties (e.g. suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?

➔ General

Contractual arrangements and the **Act of 4 April 2019** are relevant for these guarantees.

➔ Data Protection

Under the **GDPR**, the data subject has **several options** for notifying these elements. Examples include the right to rectification of inaccurate data and the right to have incomplete personal data completed, including by means of providing a supplementary statement (Art. 16). The data subject also has the right to the erasure of data (Art. 17 GDPR). Under certain circumstances, the data subject also has the right to obtain from the controller the restriction of processing (Art. 18). The data subject has the right to object to the processing of personal data, including automated individual decision-making of profiling (Art 21). The data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her (Art. 22).

➔ Consumer protection

There are also several relevant provisions found in consumer legislation. For example, Art. VI.47 **CEL** provides for a 14-day right of withdrawal, on the basis of the Consumer Protection Directive. Before a consumer is bound by a contract, the trader must also provide information in a clear and comprehensible manner, for example on the possibility of having recourse to an out-of-court complaint and redress mechanism, to which the trader is subject, and the methods for having access to it (Art. VI.45 CEL).

Art. 1649ter-1649sexies of the Civil Code also provide, on the basis of the Directive on certain aspects of the sale of consumer goods and associated guarantees, a right of recourse for the consumer as well as other rights in the event of lack of conformity. A number of remedies for consumers are also included in the Digital Content Directive and the Sales of Goods Directive in the case of non-delivery or lack of conformity.

➔ Submitting complaints

Complaints can in certain cases also be submitted to the competent authorities, such as reporting **incidents involving medical devices to the Federal Agency for Medicines and Health Products (FAMHP)**. Furthermore, a complaint can also be submitted to the **Reporting Centre** (Meldpunt) in the event of deception, fraud and scamming. Under the product safety provisions in the CEL, producers and distributors shall immediately inform the **Central Reporting Centre for Products** if they know or should know, on the basis of the information available to them and as a matter of professional judgement, that a product or service presents risks to the user that are incompatible with the general safety obligation (Art. IX.8). The **NIS Act** also provides for the possibility of reporting incidents (Art. 24-31).

➔ Sectoral legislation

The necessary processes/measures can also be included or envisaged in sectoral legislation. On can for example think of the **post-market surveillance obligations** for medical device manufacturers. The post-market surveillance plan must include provisions on, inter alia, collecting and using the available information, including feedback and complaints from users, distributors and importers.

	For applications that can adversely affect individuals, have redress by design mechanisms been put in place?

➔ General

There is a potential risk that AI systems cause harm. Examples include accidents caused by autonomous cars or a robot causing physical harm during a surgical procedure. There are **various recourse possibilities** for victims with regard to damage.

➔ Contractual and non-contractual regimes

Several aspects can be included in the **contract**. For example, contractual damages or the termination of the contract in the event of a breach.

Besides contractual liability, a victim may also make a claim under a **non-contractual liability regime**. For example, the fault-based liability of an AI producer or software developer (Art. 1382 of the Civil Code) or the liability of the custodian of a defective good (Art. 1384 of the Civil Code).

Product liability regulations are also important. According to this framework, a producer is liable for any damage caused by a defect in its product.

➔ Consumer protection

The consumer's **right of withdrawal** and the remedies/rights available in the event of non-delivery or lack of conformity have already been covered.

The Unfair Commercial Practices Directive and the relevant provisions in the CEL also provide for **redress**. Unfair commercial practices on the part of businesses towards consumers are prohibited. Member States must ensure that there are appropriate and effective means to combat unfair commercial practices, so that compliance with the Directive can be enforced in the interests of consumers (Art. 11).

Directive 2019/2161 **regarding the better enforcement and modernisation of Union consumer protection rules** was also recently adopted. Member States must ensure that remedies are available to consumers who have incurred damage due to unfair commercial practices to mitigate all the effects of such unfair commercial practices. The consumer must have access to redress and, where appropriate, a price reduction or termination of contract in a manner which is proportionate and effective.

Companies can invoke the provisions of the CEL on **misleading and comparative advertising** (Art. VI.17 et seq.) which are based on Directive 2006/114 concerning misleading and comparative advertising. The intention is to protect professionals from misleading advertising by other companies, which is considered an unfair trade practice. According to the Directive, Member States must ensure that adequate and effective means exist to combat misleading advertising and enforce the provisions on comparative advertising, in the interest of traders and competitors (Art. 5). Furthermore, the **Belgian B2B Act of 4 April 2019** on abuses of economic dependency, unfair terms and unfair market practices between companies can also be relied upon. If the rights of a consumer or a company have not been respected or if someone is the victim of deception, fraud or scamming, a procedure can be started with the **Reporting Centre** (Meldpunt)**.**

➔ Judicial Code

Legal proceedings may be instituted in accordance with the provisions of the **Judicial Code.** However, the action cannot be admitted if the plaintiff lacks the capacity and interest to file the claim. The **burden of proof** for damages caused by AI systems can be high (e.g. "defective" AI).

➔ Alternative dispute resolution procedures

In addition, **alternative dispute resolution procedures** such as arbitration or mediation may also be possible. The European Online Dispute Resolution (ODR) platform is made available by the European Commission to make online shopping safer and fairer by providing access to effective dispute resolution tools. At the EU level, a proposal for a Directive on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC has also recently been adopted.

➔ Legal actions for collective redress

In this context, the **regulations on legal actions for collective redress** are also relevant. A legal action for collective redress is admissible if several conditions are met. A class representative is allowed to introduce an action for collective redress where a group of consumers has suffered harm resulting from a company's breach of contract or violation of certain specified statutory provisions and EU regulations relating to consumer protection (Art. XVII. 37 CEL, or one of their implementing decrees). There are also a number of requirements for the plaintiff and recourse to legal action for collective redress must appear more effective than (or superior to) an individual civil action. The possibility of legal action for collective redress has also been provided for SMEs.

➔ Sectoral provisions

Certain (sectoral) legislation sometimes provides for **redress**. Under the conditions laid down in the **GDPR**, the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her (Art. 21). Moreover, the GDPR provides for a specific liability regime (Art. 82).

Where are the possible points of improvement or focus points?

	**Identifying and (where appropriate) remedying evidence issues**

The analysis shows that there are already **many redress options** available when damage is caused by AI systems. However, the problem often lies in proving the constitutive elements of the claim. It is therefore useful to further explore the **law of evidence** in an AI context. By doing so, it can be verified whether it is adapted to the reality of AI and be ensured that the victim is in no way left out in the cold.

A second component that is also involved is **digital literacy**. The question is whether the victim is actually able to effectively identify problems and gather the necessary evidence. For personalised services, the so-called *filter bubble* could pose an additional problem. People may not be aware that they are subject to personalisation and that others possibly are subject to a different kind of personalisation. The question is thus whether a potential negative influence can be recognised at all.

	**Concretising the ethical AI framework**

There needs to be a reflection on **concrete measures regarding the ethical framework** for AI in Flanders.

Policy makers in Europe are often still working at a rather abstract level of principles and the question whether new legislation should be considered. These initiatives are often too abstract to be included in the current AI-related innovation and development in Europe (or Flanders). Concrete and specific measures are therefore needed, such as appointing a Chief AI Ethics Officer in a company that develops AI systems. A number of relevant questions in this regard include what skills this person needs to have and what the job profile of such an officer looks like.

When creating a proper ethical framework consisting of concrete measures or even legislation, the 'checkbox ticking syndrome' should be remedied and be prevented at any time (cf. negative trend in personal data protection that we already highlighted above). Another possibility is that AI and data innovation processes supported by public funds need to include ethical aspects/considerations related to innovation as a selection or eligibility criterion. Ethical assessments take time and resources. They should therefore not be included as a separate section/part/heading in a project proposal but as a mandatory work package. The specific content of such a work package can/should be discussed with the various funding bodies, policy makers, NGOs promoting human rights and the Knowledge Centre Data & Society.

Furthermore, some kind of code of conduct for AI could also be considered. Such a code of conduct could set out the expected behaviour of AI developers, organisations and producers. In addition, it could be used to specify what measures are needed to ensure/simplify the auditability of the AI system, as well as possible processes/measures that may be adopted to report potential vulnerabilities, risks or biases in the AI system.

	**Focusing on certification**

Certification mechanisms for reliable AI can play an important role in the future. Policy makers need to focus more on this and further develop a framework for certification. Setting up a multidisciplinary platform with other actors would be a possibility (cf. ETAMI).

	**Knowledge dissemination**

The **dissemination of knowledge** is once again crucial. There is still a need to organise workshops providing information on the legal and ethical frameworks applicable to AI systems (cf. 'educate your leaders'). The applicable legislation could also be further clarified, especially with regard to AI itself. The question whether software is a product or when software is defective for instance deserves further attention. Designing a contractual template in the context of AI and standard provisions on the division of liability/responsibility can be useful in the AI supply chain as well.

What tools can be used to fulfil the ethical requirement?

- Data Ethics Guide aims to highlight the issues surrounding ethics and make them manageable for companies working with AI technology. Various concepts are introduced on how ethics can be looked at within a company. This is done through questions that help determine what the current ethical situation is.

- AI systems Ethics Self-Assessment Tool
- Principles for Accountable Algorithms and Social Impact Statement for Algorithms
- Data Ethics Decision Aid (DEDA)
- SDoC or Supplier's Declarations of Conformity
- Ethics framework of Machine Intelligence Garage
- Data Collection Bias Assessment form
- AI Explainability 360
- Ethical OS tool kit
- Data Ethics Canvas
- Artificial Intelligence Impact Assessment
- Aequitas
- Building an Algorithm Tool

EVALUATION

For each ethical requirement, several potential shortcomings have already been identified and a number of recommendations were formulated. In this section, we conclude by giving a brief evaluation of the relationship between the ethical guidelines and the legal framework. We identify some gaps that policymakers could focus on and also give an overview of elements that (may) need further consideration. One thing is immediately clear: anyone who wants to 'link' the ethical requirements to the legal framework must have a very broad knowledge of almost all legal domains. As such, in an age of increasing specialisation, an all-round legal knowledge would certainly seem to offer an advantage with regard to developing reliable AI systems. Moreover, the ethical requirements also require a specific mindset that cannot always be translated into legislation (e.g. environmental, social, sustainability).

➔ **Requirement 1 (Human agency and oversight)**

This ethical requirement is already reflected in many and various legal requirements. It is therefore primarily a question of clarifying and interlinking existing rules. For example, the current regulations are still not intended for systems that can 'learn' or take unforeseen actions by the designer/user. It is thus recommended to (i) adapt/supplement any relevant legal requirements or technical standards so that they can take into account the autonomous nature of AI systems and (ii) examine to what extent the scope of certain legislation needs to be adapted. On the other hand, this requirement also relates to concepts that are legally less clear-cut such as 'addiction' and 'attachment', and which are not yet currently covered in law. As such, the question is whether and to what extent the (European or Belgian) legislator can or should act to convert these concepts into legal requirements, not only for AI systems but also for other products.

➔ **Requirement 2 (Technical robustness and safety)**

In terms of technical robustness and safety, there are already several legal frameworks requiring that adequate safety be envisaged. Digitalisation – and especially the increasing use of AI systems – are 'testing'/'stretching' these rules at several levels. Many existing regulations cover tangible and movable products (e.g. product safety and product liability). The cybersecurity regime is primarily focused on preventing attacks by hackers. Additional interdisciplinary cooperation and research seems thus necessary to examine whether and how this applicable legislation and/or regulations need to be refined in an AI context.

➔ **Requirement 3 (Privacy and data governance)**

With regard to privacy and data governance, there is already an extensive legal framework in place based on data protection law. Nevertheless, effectively applying it often proves problematic. It is thus difficult to achieve a workable privacy culture in practice. This can presumably be resolved by increasing awareness and providing a clearer explanation of the impact and application of data protection law in an AI context. Better enforcement also seems inevitable in this regard. One problem is that only processors and controllers are subject to the GDPR. As such, developers and providers are strictly speaking not required to make their (AI) products compliant with inter alia data protection requirements by design and by default. Regulatory initiatives in this respect can only be useful if they are taken at least at the EU level.

➔ **Requirement 4 (Transparency)**

There is already much legislation on transparency and information obligations, especially with regard to consumers. As such, additional legislation does not seem necessary although the existing legislation/legal concepts could (in the future) be slightly modified and/or clarified in some areas. With some small additions and modifications specifically targeting AI systems (e.g. concerning input data and the functioning of AI systems),

much can already be achieved in terms of compliance with the ALTAI list. One alternative would be to adopt additional guidelines specifically for AI systems (cf. 'AI leaflet').

➔ Requirement 5 (Diversity, non-discrimination and fairness)

In the area of non-discrimination, there is already a comprehensive legal framework. However, it seems that possible new forms of discrimination may not be covered by the existing legislation. As such, it must be examined whether this can be resolved by supplementing existing legislation. On the other hand, stakeholder participation in an AI context is not yet addressed in specific legislation. An example of such participation can already be found in the context of the GDPR. For AI systems, it must be examined what such regulation should look like.

➔ Requirement 6 (Societal and Environmental Well-being)

As regards environmental awareness, there is already an extensive legal framework that focuses on specific risks that can occur. There are also standards to make the development of products more sustainable. As these are primarily focused on hardware, it is useful to consider whether such requirements should also apply to software development. It also follows from this requirement that relevant actors need to be adequately informed about the environmental and societal impact when developing and using AI systems. Regarding the impact of AI systems on democratic values, it appears to be primarily a question of awareness. In terms of labour and social matters, it is necessary that the general public learns to use and deal with AI systems without imposing excessive (administrative) costs on businesses. Legal mechanisms are already in place for consulting and informing employees. Nonetheless, more attention can once again be given to specific training programs, especially to developing and maintaining digital skills to handle AI systems. Envisaging further vocational training is therefore crucial.

➔ Requirement 7 (Accountability)

For this requirement, a number of starting points have already been found in existing law. If harm is caused by AI systems, there are several options for victims to obtain compensation. Nonetheless, additional refinements to the legislative framework are needed, for example with regards to certain legal concepts (e.g. software qualification) or procedural aspects (e.g. burden of proof on victims).

SELECTED BIBLIOGRAPHY

Alexander J. Wulf & Ognyan Seizov, "Artificial Intelligence and Transparency. A Blueprint for Improving the Regulation of AI Applications in the EU", European Business Law Review, 31, 2020 (4), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906460

Andrea Bertolini, "Artificial Intelligence and Civil Liability", JURI committee, July 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3921216

David Leslie, "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector", June 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403301

Expert Group on Liability and New Technologies, New Technologies Formation, "Liability for Artificial Intelligence and Other Emerging Digital Technologies", November 2019, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf

Frederik Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision-making", COE, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

Gerhard Wagner, "Liability for Artificial Intelligence: A Proposal of the European Parliament", July 2014, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886294

Giovanni Sartor, "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", European Parliamentary Research Service, 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

Jacob Turner, *Robot Rules: Regulating Artificial Intelligence*, Springer, 2018, 377 p.

Jan De Bruyne & Cedric Vanleenhove, *Artificial Intelligence and the Law*, Intersentia, 2021, 520 p.

Josh Cowls, Andreas Tsamados, Mariarosaria Taddeo & Luciano Floridi, "The AI Gambit — Leveraging Artificial Intelligence to Combat Climate Change: Opportunities, Challenges, and Recommendations", March 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3804983

Karen Yeung, Andrew Howes, Ganna Pogrebna, "AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3435011

Kenniscentrum Data & Maatschappij, "Artificiële intelligentie en gegevensbescherming: een verkennende gids", 2020, https://data-en-maatschappij.ai/uploads/publications/20200602_Rapport-AI-GDPR_aug2020.pdf

Luciano Floridi & Josh Cowls, "A Unified Framework of Five Principles for AI in Society", 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3831321

Margot E. Kaminski, "Understanding Transparency in Algorithmic Accountability", U of Colorado Law Legal Studies Research Paper No. 20-34, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3622657

Martin Ebers & Susana Navas (eds), *Algorithms and Law*, Cambridge University Press, 2020, 319 p.

Mark Coeckelbergh, *AI Ethics*, MIT Press, 2020, 248 p.

Markus Dirk Dubber, Frank Pasquale & Sunit Das, *The Oxford Handbook of Ethics of AI*, Oxford University Press, 2020, 896 p.

Marcelo Corrales, Mark Fenwick, Nikolaus Forgó, *Robotics, AI and the Future of Law*, Springer, 2018, 237 p.

Matt Hervey & Matthew Lavy, *The Law of Artificial Intelligence*, Sweet & Maxwell, 2021, 588 p.

Miles Brundage, *The Malicious Use of Artificial Intelligence: Forescasting, Prevention and Mitigation*, Future of Humanity Institute, 2018.

Philipp Hacker, "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973

Richard Warner & Robert Sloan, "Making Artificial Intelligence Transparent: Fairness and the Problem of Proxy Variables", January 11, 2021, https://ssrn.com/abstract=3764131

Sandra Wachter, Brent Mittelstadt, Chris Russell, "Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI", Computer Law & Security Review 41 (2021): 105567, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922

Sandra Wachter & Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", Columbia Business Law Review, 2019 (2), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829

Sebastian Lohsse, *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 2019, 352 p.

Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, Springer Nature, 4 nov. 2019, 127 p.