Standard for Trustworthy Artificial Intelligence

How to manage AI risk?

A guide for Norwegian businesses

A collaboration between EY & Langsikt

V.01 - Updated as of June 11, 2024

Summary

i. The Standard in brief

This document (the "Standard") provides public and private enterprises with an overview of the risks of artificial intelligence ("AI") and what can be done to address those risks.

The standard has been prepared jointly by the think tank Langsikt and Ernst & Young Advokatfirma AS ("EY").[1] Langsikt is a politically independent think tank that sheds light on the most important and neglected problems of our time. EY is a leading law firm in technology, among other things, and is part of the global EY network.

The standard is intended as a tool for Norwegian enterprises, private and public, to help enterprisesthem manage the risks associated with artificial intelligence.[2] The Standard is intended to serve as a practical manual for navigating the complexities of the technology, regulations, and ethical considerations. It has been updated in line with the date on the first page and will be kept updated continuously in line with technological, economic and legal developments.

ii. How to use this document

The standard is structured according to the primary choices an enterprise faces in its encounter with AI and the factors the enterprise should consider in this regard. It provides guidance in the face of three key questions: What are the risks of AI, what are the sources of risk and what should the enterprise do to manage those risks?

The standard describes considerations that must be taken into account in the various parts of an AI system's life cycle .

1. Data collection
2. Data processing
3. Training, fine tuning and/or adaptation of use
4. Internal or external use
5. Maintaining and updating the AI model

It will be helpful to read the document in its entirety, but the reader can also jump to the sections that are most relevant. How the Standard is used will depend on the type of business and the role of the reader, see Appendix 1: Relevant user groups – who is the Standard aimed at?

The Standard contains structured methodology, guidelines for risk management and recommendations for trustworthy use of AI. It promotes the need for governance structures that can cope with unforeseen events and maintain trust in the business. See Appendix 2: Key Terms and Definitions relevant to understanding this Standard.

The Standard is not intended to be exhaustive and does not provide advice for specific situations, but the guide is intended to provide enterprises with guidance in thinking about how risks associated with AI can be managed. The enterprise must assess, for itself, its need for specific legal advice within its given situation, and EY and Langsikt do not take responsibility for how the Standard is

---

[1] From Long-term, Jacob Wulff would (advisor), Aksel Braanen Sterri (technical manager) and Jakob Graabak (senior advisor) have contributed to the Standard. Long-term also thanks Julia Graham and Hanna Malm for their contributions. From EY, Mads Ribe (Associate Partner/Lawyer and Head of AI, EY Tax & Law), Andreas Bjørnebye (Associate Partner/Lawyer, EY Law) and Benjamin Green (Associate, EY Law) have contributed.

[2] Ethical and legal risks, along with a lack of expertise, are mentioned in a survey, conducted by EY, from September 2023 as the most significant barriers to the use and scaling of generative AI for Norwegian businesses. The standard therefore places extra emphasis on these considerations.

used and are not responsible for any errors or omissions either in the Standard or resulting from use of the Standard in specific situations.

iii. Trustworthy AI is a competitive advantage

Trust is a pillar of Norwegian society. Our aim is for the Standard to help Norwegian businesses maintain and manage parts of this trust by identifying and implementing measures to limit risks in the development and use of AI.

A systematic approach to AI and risk could give the enterprise a competitive advantage. This will make it easier for the enterprise to meet regulatory requirements, promote responsible innovation, and manage risks related to AI in the face of future challenges, which in the worst case could shake the enterprise's existence.

iv. Four principles for trustworthy AI

Businesses should consider the following basic approach when developing and using AI:

1. Appropriate: Make sure to use AI to achieve overall goals for the business and that AI is only used where the benefits of its use outweigh the disadvantages.
2. Risk management: The business should identify risks and have a systematic approach to reducing risk in line with the expected benefits.
3. Considerations: Take ethical and legal considerations into account in the enterprise's development and use of AI and ensure that its use is in line with regulations, industry standards and reasonable expectations from users and society at large.
4. Holistic: Make sure that all functions in the business and all users experience the positive effects of AI.

TABLE OF CONTENTS

Appendices:

1 Should the business use artificial intelligence?

1.1 What is artificial intelligence?

Article 3 of the AI Act defines an AI system as follows:

> 'AI system' means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.[3]

A distinction is often made between two types of AI:

1. Generative AI is probably best known through ChatGPT, and gets its name from the system's ability to generate new text, images, audio, and videos.[4] The most general generative models increasingly also constitute a new infrastructure, called foundation models. Foundation models have good general skills, which can then be specialized as needed.

2. Narrow AI is adapted to a narrow area of use, where it is often very good. For example, such systems can be used for diagnostic imaging, play certain games or calculate the 3D shapes of proteins based on a genetic code.

Both types of AI are based on machine learning. Previously, smart rules were programmed into the AI system. Today, we let the model create its own rules by rewarding it for coming up with the result the developers want during training. For example, a model can learn to detect cancerous tumors by practicing images with and without tumors, and provide rewards when the model guesses correctly. To train such models, data is needed that the model can train on and the computing power to perform the computational operations as a basis for the training.

1.2 Strategic considerations when using AI

The overall most important things the enterprise should consider in relation to artificial intelligence are:

1. Which specific needs can the enterprise satisfy by developing or using AI? For example, automating existing work tasks or using AI to solve problems that have not been able to be solved with existing expertise.

2. Is there a need for the enterprise to build AI competency, which will create long term benefits?

3. What is the need for the enterprise to adapt its business model or business strategy according to technological development?

The use of AI must be adapted to the individual enterprise and the tasks it performs. This must be assessed on an ongoing basis according to which AI solutions are available and how AI changes the market in which one operates. Since technology and the supply of new models are changing rapidly, the enterprise should have an overview of the current situation and a plan for different scenarios going forward. See Appendix 3: Strategy checklist for strategic choices.

---

[3] KI-forordningens artikkel 3.

[4] The development in the use of AI, and generative artificial intelligence in particular, has accelerated since the American company OpenAI launched ChatGPT in November 2022. Since then, several similar solutions have come on the market in the form of Google Gemini, Anthropic Claud, Meta Llama and others.

2 Risk management

2.1 Five types of business risk

Risk is the product of the potential for damage in the event of an adverse event and the likelihood of the incident occurring. The greater the probability that the injury will occur and the more severe the injury, the greater the risk.

Overall, there are five types of risk the enterprise faces:

1. Legal risk: risk of breach of legal obligations, both current and future, with civil or criminal sanctions if the business breaches the obligations.
2. Ethical risk: the risk of acting in violation of ethical norms.
3. Value risk: the risk that the business does not act in line with the enterprise's values.
4. Task risk: the risk that the enterprise does not perform the enterprise's tasks in a satisfactory manner.
5. Disruption risk: the risk that technological and societal changes undermine the enterprise's raison d'être.

Taken together and individually, these business risks can threaten the enterprise's ability to achieve its goals or exercise its functions.[5] Exposing yourself to the different types of risks can also cause damage to the enterprise's reputation. The Standard will have the most to say about legal risk, ethical risk and task-specific risk.

Note that the five business risks are neutrally formulated - they are not specific to AI. This reflects the fact that AI can both increase and reduce business risk.

AI can be used in ways that are contrary to legislation and values and that weaken the quality of the product the enterprise delivers but can also be used to deliver better products and better comply with laws, regulations and values.

AI can provide new opportunities that were not previously available to the business, but also undermine the need for the core business. If the enterprise is unable to exploit the opportunities in AI in a way that makes it relevant and efficient, this could entail a significant competitive disadvantage.

2.2 How should businesses take risk into account?

In the face of risk, it is essential that the enterprise:

1. Obtains an overview and information about the risk and the sources of risk
2. Acquires relevant expertise to manage risk
3. Allocates responsibility for managing AI
4. Taking action to manage or minimize risk
5. Identifies weaknesses or shortcomings in the measures, evaluates and ensures continuous improvement in line with use and technological development
6. Puts in place good governance systems for 1, 4 and 5 (governance) that reduce risk and create trust among owners, users, employees and society (which in turn reduces risk)

[5] All four categories are elaborated further down in the document and can be broken down into subcategories. For example, one can talk about model risk and other more specific risks.

7. Continuously monitors the enterprise's management systems.[6]

2.3 Take reasonable steps to reduce risk

Risks are generally something businesses live with. Zero risk is not practically possible, neither in the business sector, the public sector nor in society at large. Risks associated with the development or use of AI must therefore not be eliminated but managed responsibly.

Owners, employees, creditors, customers and the rest of society will expect the business to take reasonable steps to reduce the risk of various incidents occurring and to manage the situation if a risk occurs.

What is considered reasonable will largely depend on a cost-benefit analysis: How much will a measure reduce the risk, at what cost to the business and society?

Some measures cost more than it tastes in relation to the reduction in risk achieved. As a rule, such disproportional measures should not be implemented. Among proportional measures, the most effective ones should be adopted.

2.4 Principles for responsible and ethical use of AI

To prevent ethical risk, the enterprise should act in line with principles for ethical and responsible use of AI.

There is a large overlap between the ethical and the legal. The rules of law are often based on ethical considerations, such as justice, but the overlap is not perfect: not everything that is immoral is illegal and not everything that is illegal is immoral. In the list of the most relevant ethical considerations, there will nevertheless be considerable overlap with legal rules, which will require legal, as well as ethical, assessments.[7]

The most relevant ethical considerations are:

1	*Fairness*	Everyone is entitled to equal rights, opportunities, and fair treatment. The enterprise should not discriminate without sufficient good reason and should ensure that it does not use AI in a way that is unfair.
2	*Prevent damage*	Everyone should do what is reasonable to prevent harm to others as a result of their actions and choices. This also applies to enterprises, which should prevent users or third parties from being harmed by AI the enterprise uses.
3	*Self-determination*	Every adult has the right to self-determination. The AI systems the enterprise uses should strengthen, not weaken, the ability of users and other affected parties to make good choices. AI systems, for example, should not exploit people's vulnerabilities.
4	*Privacy*	Everyone has a private life that they should have in peace. Where the enterprise uses AI, it should be ensured that the privacy of both users and third parties is respected.[8]
5	*Good reasons*	Decisions that affect others must be justified. The enterprise must therefore be able to justify its use of AI systems to users, authorities and others affected by the system.

[6] See the complete checklist in Annex III
[7] Several of the items in the list overlap with the EU's Ethics by Design and Ethics of Use Approaches for Artificial Intelligence of 25.11.2021:.
[8] See the Norwegian Data Protection Authority's overview of the privacy principles.

6	*Understanding*	Good reasons require understanding from others. The functioning and decisions of the AI system must be understandable by users, authorities and others who are affected by the system.[9]
7	*Responsibility*	Everyone – both natural and legal persons – is responsible for their own choices and actions. The enterprise must take responsibility for decisions made by the AI system and must therefore understand, monitor and control the design and operation of AI-based systems.

2.5 Legal risk: "Legal AI"

Responsible AI involves a conscious relationship to the ethical considerations that apply and the extent to which the context entails a risk of violation of legal rules. As mentioned above, there is a large overlap between the ethical and the legal, where the legal rules often operationalize ethical considerations, such as in the area of privacy or liability for damage. An understanding of the basic principles behind regulations that are relevant to AI (even those that have not yet been implemented in Norway) can help reduce legal risk and strengthen the enterprise's position.[10]

Norway has, to a large extent, a technology-neutral legal framework: actions are regulated regardless of the technology or method used.[11] Even though AI is not directly regulated today (June 2024), there will still be legislation that applies to the development and use of AI.[12]

In addition to the current legislation, it is the EU's AI Act that is most relevant to relate for Norwegian enterprises. The AI Act breaks with the tradition of technology-neutral legislation and directly regulates AI. The Norwegian authorities have expressed that the regulation will be implemented in Norwegian law.

In addition to existing or future legislation, it is useful for the enterprise to investigate whether there are relevant standards that can be useful for either certifying its own model or for checking the certification of a developed model the enterprise plans to use.[13] The enterprise must itself assess what is relevant for its development or use of AI, such as taking into account industry, geographical location, risk profile and level of knowledge.

A enterprise that operates in several legal areas must be aware that different jurisdictions have different regulations. We have therefore included an overview of regulatory trends internationally in Appendix 4: Overview of International Regulations.

---

[9] An explanation of how the system relies on pattern recognition (or similar logic) to adapt to and perform many different tasks, could in itself be considered an adaquate explanation in some scenarios.
[10] An understanding of the AI Act will increase awareness of AI risks, prepare the business for the transition to the requirements that apply under the Act, build trust with customers about the development and use of AI, and increase credibility with authorities. Good preparation will make the transition easier and potentially provide a competitive advantage when the regulation is implemented. It can also help the company anticipate governance needs and compliance requirements that may be relevant to the company's specific development and use of AI.
[11] For example, a ban on theft, regardless of whether the theft is digital or physical and regardless of the method or technology used.
[12] This includes intellectual property rights (such as protection of intellectual property under the Copyright Act), data protection legislation (e.g. protection against unlawful use of personal data), the ICT Regulations for security in IT systems, anti-discrimination legislation (such as protection against discriminatory behavior). Furthermore, contractual obligations may also affect the use of technology (such as outsourcing agreements for IT services).
13 See Standards Norway's page on AI standards.

2.5.1 Classification of AI systems under the AI Act

The EU's AI Act regulates AI according to risk, roles and area of use. The enterprise must have an active approach to its role it has in different situations where AI is involved, and at the same time be able to categorise its use of AI according to the various risk categories.[14]

The risk categories are as follows:

Category	Regulation	Description	Examples - AI systems which:
Unacceptable risk	Prohibition	Systems that pose unacceptable risks to the fundamental rights of people, or that can expose them to physical or psychological harm.	• Exploit people's vulnerabilities or manipulates a specific group of people (e.g., children, the elderly, disabilities) and circumvents the users' will • Perform scoring for socially desirable behaviors • Interpret employee emotions in the workplace • Perform biometric categorization to interpret sensitive data • Perform untargeted facial recognition stored in databases • Perform predictive monitoring of individuals to prevent crimes that may happen
High risk	Legal but strict terms	Systems that pose a significant risk of harm to health, safety or fundamental rights. Most of the obligations in the AI Act apply to such systems, divided between suppliers and users.	• Are used as safety components in other products and systems (e.g., in medical devices, automobiles, machinery) • Pose a significant risk of harm to health, safety or fundamental rights, including biometric ID, management and operation of critical infrastructure, education, employment, access to public services, certain forms of government exercise (e.g. police, asylum), democratic processes (electoral systems), etc.
Limited risk	Legal, but limited terms	Limited transparency requirements	• Interact with humans; these must clearly inform that an AI system is being used (e.g., chatbots) or that AI-generated material (e.g. deepfakes) is being viewed.

[14] This is to take into account the company's current requirements under existing legislation, both general and sector-specific requirements (e.g. data protection legislation, intellectual property legislation, marketing legislation, data security and ICT regulations, the EU Machinery Directive, etc.), the company's contractual obligations (to the extent that the use of AI affects the company's contractual rights and obligations, e.g. related to transparency about and use of the result/outcome of AI use in specific situations) or future requirements under the AI Regulation and other relevant legislation such as the AI Liability Directive and the updated Product Liability Directive (both from the EU).

Category	Regulation	Description	Examples - AI systems which:
			• Certain exceptions exist for artistic works, or content that is obviously AI-generated.
Minimal risk	Legal, but few or no terms	All other AI systems that do not fit into the other categories	• Typically, work processes increase efficiency without affecting users' rights (e.g. photo editing, product recommendation, spam filter, translation tools, etc.)

Obligations under the AI Act also depend on the role of the enterprise in relation to the development or use of AI.

Role	Description	Demand
Provider	Natural or legal person, public authority, organization or other entity that has developed an AI system that is rolled out on the market, or to put into operation under its own name or trademark, either for payment or free of charge.	Requirements for providers of high-risk AI systems include: • Establishment and maintenance of appropriate risk and quality management systems. • Effective data management. • Human oversight. • Compliance with applicable standards (e.g. cyber-security). • Registration of the system in the EU database, including for critical infrastructure AI which is included in a non-public section of the database.
Deployer	A natural or legal person, public authority, organization or other entity that uses an AI system under its authority.	Requirements for deployers of high-risk AI systems include: • Fundamental Rights Impact Assessment (FRIA) prior to deployment of the AI system, if the user: – Is a public body or private entity that provides public services – Provides essential private services, including creditworthiness assessment, risk assessment and pricing in relation to life and health insurance. • Human supervision of people with appropriate training and competence. • Ensuring that prompts/input data into the system is relevant to the system's use. • Suspension of use of the system if it poses a risk at national level. • Informing the AI system supplier of any serious incidents. • Compliance with GDPR obligations to perform a privacy impact assessment. • Verification that the AI system complies with the AI Act and that all relevant documentation has been provided.

Role	Description	Demand
		• Informing people that they may be subject to the use of high-risk AI.
Distributor	Any natural or legal person in the supply chain, who is not a supplier or importer, but who makes an AI system available on the EU market.	Requirements for distributors when deploying high-risk AI systems include: • Check that the system has the necessary CE marking, with a copy of the EU declaration of conformity, etc. • Examine available information and assess whether the AI system is in line with the AI Act. If it is not in line with the Regulation, corrective measures must be implemented, or a recall must be carried out. • Assess whether the AI system poses a risk based on available information and cooperate with relevant authorities to reduce the risk. • Provide all information and documentation about the system, upon reasoned request from the relevant authorities.
Importer	Any natural or legal person located or established in the European Union who places on the market an AI system bearing the name or trademark of a natural or legal person established outside the European Union.	Requirements for importers when deploying high-risk AI systems include: • Ensure that the supplier of the AI system has carried out the necessary procedures and technical documentation required. • Check that the system has the necessary CE marking, with a copy of the EU declaration of conformity, etc. • Do not place the system on the market if there is reason to believe that the system violates the AI Act, is counterfeit, etc. • Provide all information and documentation about the system, upon reasoned request from the relevant authorities. • Work with relevant authorities to mitigate risks identified by the system.

The requirements imposed on the enterprise and the level of fines in the event of a breach of the AI Act will depend on the role and classification of the AI system in question. For most enterprises, it is deployer that will be most relevant. The choice of model and risk classification of the relevant AI solution will be essential for the enterprise's risk profile and framework for managing such risk (governance).

2.6 Risks associated with model selection

Overall, a enterprise that wants to use AI has three options: 1) Develop its own AI model, 2) acquire and use existing models, 3) adapt existing models. We will go into more details about the three possibilities in the rest of the document.

If the enterprise wants to develop its own AI model or conduct extensive adaptation of an existing model, the enterprise must consider the risk associated with access to training data, the risk associated with training the model, and the costs and risks associated with using the model.

If the enterprise wants to use an existing model, it is primarily questions about use that apply. In all cases, both existing regulations and the AI Act will have to be taken into account.

Model selection checklist:

1	Understanding the problem	The enterprise must have a clear understanding of the problem and a strategy for what AI will solve. Is the model well adapted to the problem AI is supposed to solve in the business?
2	Performance versus explainability	Some AI models can provide high performance, but are often "black boxes" with low explainability. In some cases, it may be necessary to sacrifice some performance to achieve greater explainability and trust, especially in regulated industries.
3	Resources and expertise	Does the enterprise have the necessary resources and expertise to develop and maintain the chosen AI model? If not, you may want to choose off-the-shelf products.
4	Timeframe and costs	Implementing technology takes time and can be resource-intensive. Rapid and cost-effective implementation of an AI model must weigh up the importance AI could have for the automation of tasks and strategic significance for the business through a more holistic and tailored approach.
5	Scalability	The choice of AI model must be scalable across the enterprise's activities, both now and in the near future. Furthermore, the model should be under constant development and improvement.
6	Security and privacy	The business must ensure that the AI model complies with applicable security standards and privacy legislation, especially when it comes to the processing of sensitive data.
7	System integration	The AI model must be able to be integrated with the enterprise's existing IT systems and processes for optimal use. Integration will depend on the provider's standard integrations and the enterprise's existing IT infrastructure.
8	Influence and ethics	The business must consider the expected impact of the AI model on customers, employees, and society.

The checklist should inform whether you 1) develop your own AI model, 2) acquire and use existing models, or 3) adapt existing models, and whether you choose open or closed models. Information about customizing existing models and open and closed solutions can be found in ustomizing foundation models.

3 Training data

Questions related to the risk of training data are primarily relevant for two types of use:

1. Where the enterprise develops a new AI model or fine-tunes existing AI models with tailored training data specific to the enterprise's needs. Developing or fine-tuning an AI model requires the enterprise to acquire, manage, and make available data for training the model.
2. Where the enterprise procures an external AI model that is trained on unknown or all-encompassing data (typically by data scraping from the internet). Training a model on unknown or all-encompassing data can result in encroachment on the (intellectual property) rights of others.

When collecting, handling and using data, the enterprise must take three types of considerations:

i. Legal: Comply with applicable regulations
ii. Ethical: Take adequate steps to ensure that data collection is conducted ethically
iii. Accuracy: Quality check the data, so that the model is accurate enough to achieve the purpose of the business.

As we will see, the different considerations can come into conflict with each other.

3.1 Legal risk

Before the enterprise collects and/or makes data available to the model for training purposes, the enterprise should consider the data's legally permissible uses. This may result from various legal bases, typically an agreement with the owner or legislation. The importance of assessing the lawful uses of the data can be illustrated by the following: If the enterprise wishes to train a model to be used in one area and the data can only be legally used in another area, the enterprise has a legal challenge in the use of the data.

The enterprise must ensure a legal assessment of the data in the specific case. However, there are several regulations that regulate different data.

3.1.1 Personal data

Personal data from citizens in the EU/EEA is regulated under the General Data Protection Regulation (GDPR) and is implemented in Norwegian law under the Personal Data Act. Personal data is a broad term and includes any information about an identified or identifiable natural person. Any use of personal data must have a legal basis.

Certain data (sensitive personal data, such as information about a person's ethnic origin, political views, religion/philosophical beliefs, genetic factors, biometric data, health, sexual relations and sexual orientation) are, as a general rule, prohibited to use. The GDPR regulates a number of aspects of the collection and use of personal data, and a thorough understanding of the GDPR is important when using such data.

3.1.2 The Data Governance Act

Some data is subject to the EU's Data Governance Act (“DGA”). The DGA is intended to facilitate the accessibility, collection, sharing and use of data between the business sector and the public sector, such as data generated by the Internet of Things. This may affect various contractual terms and the interoperability of data.

3.1.3 Intellectual property rights

Some data may be protected by intellectual property rights, such as copyright/database protection and trade secret protection under the EU Trade Secrets Directive, implemented in Norwegian law by the Trade Secrets Act. As a general rule, one should assume that data which forms the basis of an enterprise's business model, or which may be sensitive to the enterprise, is not free to use without further clarification. In the case of so-called "scraping" of data from the internet, the file "robots.txt" can provide valuable information about the rights to the data, but this does not guarantee that the data is for free use.

3.1.4 License terms

Although data may be subject to rights, they may be available for use under different licence terms, which either follow general standards or which must be negotiated specifically. Examples of general license terms are Creative Commons (CC), MIT and NLOD. Some license terms allow the data to be used for training an AI model, but this must be considered on a case by case basis.

3.1.5 Discrimination

Correct information is not sufficient to ensure that the AI model is in line with the regulations. If data documenting discrimination or marginalisation of certain groups or individuals are used for training an AI model, the discrimination can be reinforced or at least continued by the model. This may conflict with the legislation's protection against discrimination.

3.2 Ethical risk

In addition to the various legal considerations that apply to data for training an AI model, there are also various ethical considerations that apply. There may be some overlap between the legal and ethical considerations. For example, it is important to ensure that the person to whom the personal data relates has understood what the data will be used for, so that the consent is informed and that the data is used in line with the data subject's reasonable expectations.

There are also problems related to biases and weaknesses in data, where data documenting discrimination or marginalisation can perpetuate or reinforce such trends in the AI model. The enterprise has a responsibility to counteract this.

3.3 Task risk

3.3.1 Representativeness

Furthermore, it should be considered whether the data used is representative of what the model will be used for. If the data on which the model is trained is from a different population than the one to which the model is to be applied, the enterprise risks imprecise or incorrect results from the model. However, the fact that the data is not representative does not mean that it should not be used. Expected differences between training and use should be mapped so that this can be taken into account in training, testing, use, and communication of the model.

3.3.2 Data poisoning

Another consideration that should be considered for a correct use of the data is whether the training data may be exposed to so-called "poisoning" (training data poisoning). For example, vulnerabilities in a spam filter can be created and then incorrectly categorized examples can be entered into the training dataset. Such weaknesses can be exploited.

There are several techniques for introducing traps and errors into datasets. Data scraped from the web may therefore have reduced value. This should always be considered and is often related to the

question of whether the data is subject to copyright protection (see above about legal considerations).

4 Training method

If an AI model is to be useful to the business and not have undesirable and harmful results, it must be accurate. If the model is to be accurate, the AI model's architecture, target function, and training data, *combined*, must be suitable. One must create a theory about what kind of architecture and target function can best achieve the enterprise's goals given the training data one has.

4.1 The architecture of the AI model

The architecture is designed for the AI model. It describes the characteristics of an AI model that do not change during training. The choice of architecture determines how good the model can be, how much computing power and data is required to train the model, and what type of errors the model is likely to make.

For example, some image analysis models have a built-in assumption that pixels far away from an area are irrelevant to understanding what is happening in that area, and that the same types of relationships between neighboring pixels apply throughout the image. Other image models always look at the whole image in context.

Every architecture has implicit assumptions about the shape of the training data and the problem it is supposed to solve.

A good choice of architecture is therefore adapted to the characteristics of the data and the goal of the AI model. The AI model can then become good at what is important and bad at irrelevant skills by using the training data, without requiring an unnecessary amount of computational power.

4.2 The objective function of the AI model

An objective function is a metric that defines the goal of an AI model. The AI model is trained to either maximize or minimize this number. Large language models such as ChatGPT, for example, are trained to minimize the difference between what it thinks is the next word in a number of texts and what is actually the next word. The fact that an objective function requires you to reduce overall goals to a single metric makes it difficult to ensure that the objective function, and thus the AI model, manifests the enterprise's wishes.

The selected objective function can lead to under-prioritization of some scenarios; i.e., it works well in most cases, but fails disproportionately in some situations. Applied to individuals, it can lead to indirect discrimination (disparate impact). A lack of precision will also entail a task risk (e.g., if the model is wrong in decisions that are particularly important to the business).

In addition to ensuring that the model is accurate, the enterprise must also consider that all data and labels made available to the model are in line with applicable legislation and ethical considerations. Given the purpose of the model, what are legitimate decision variables? For example, is it in line with anti-discrimination legislation and ethical principles to use data on race to train the model?

4.3 The need for testing

In order to improve the model and form the basis for appropriate use, good and ongoing testing of the model is essential. Test your theory for choice of architecture and target function, does the model work as you thought it would?

1. Map precision for different uses: Check if the biases you mapped in the dataset have propagated to the model, and test for known failure modes for the architecture you chose.

If one is to apply the algorithm to personal data, one must test for discrimination based on characteristics such as gender, ethnicity, disability, pregnancy and leave.[15]

2. Explainability: Try to understand how the model operates, for example by principal component analysis. Is it possible to explain how the model arrives at the result? If the model is used to make decisions that are relevant to people, it may be illegal, unethical, or commercially risky to make decisions based on the model's results without understanding why or how the model made its assessment. Therefore, use the training and testing phase to understand the model.
3. "Red teaming": Try to actively break down your own model with targeted attacks of possible weaknesses. Here, particular focus should be placed on identifying the weaknesses in the data basis and known failure modes of similar models.
4. Standards: Should the model be certified according to a relevant standard? Standards can assist in defining the necessary testing approach, make it safer for customers to use an AI model, and contribute to more transparency and better industry practices.[16]

4.4 Customization of foundation models

A foundation model is a type of AI model that has undergone extensive general training and can therefore easily be used or adapted for different purposes. The most well-known are large generative AI models[17] that can handle text, images, audio and video. There are also more domain-specific foundation models.

If the enterprise is based on a foundation model, one should obtain an overview of how the model is trained. The model card provides useful information about the model, training basis, precision and recommended area of application.[18] Customization can be done mainly in three ways: fine training, few-shot prompting, and data access.

4.4.1 Fine training

Fine-tuning a foundation model involves training the model on more data to specialize the model in exactly what the enterprise wants to use it for. Since the foundation model has trained so much in advance, it often just needs a little extra training to hone in on the purpose of the business. Fine tuning can therefore provide good models for little data and computing power. Often, one of the large parts of the architecture in the foundation model locks before the fine training, in order to preserve the knowledge the model already has.

If the enterprise fine-tunes the model, the same considerations apply as in section 4.3, but seen in the context of the information about the foundation model. The target function and the training data for the fine training and the architecture, precision and training basis of the foundation model must together be suitable for an accurate model.

4.4.2 Few-shot prompting

An even easier way to adapt a foundation model is so-called "few-shot prompting". Before a user request gets to the base model, several examples of requests and good answers are automatically pasted. The examples allow the model to understand what type of response it should give before it then tries to give such a response to the request. Just a couple of examples can make the model much better. No training is required, just a set of sample requests and responses.

---

[15] For testing discrimination, see LDO.
[16] See Standards Norway's page on AI standards.
[17] See footnote 4 for examples.
[18] See overview from Hugging Face.

Few-shot prompting gives the enterprise increased control over the type of response the model gives. But you guide, not override, the underlying model. The enterprise should therefore be aware that it will continue to behave as the foundation model has been trained to. The examples used should be good, since the model will use each example for every request.

4.4.3 Data access

Giving a foundation model data access is useful for incorporating the enterprise's own data and knowledge into an otherwise general model. This is often called RAG ("Retrieval-Augmented Generation"). It works by a search engine looking up the enterprise's own database based on the user's prompt and pasting relevant information together with the prompt. The prompt that arrives at the foundation model thus includes both the user request and the knowledge required to respond to the request. The model only needs to respond.

With such data access, it becomes important to check that the search function works well. It should only have access to data you want to share with the user and should have a reliable lookup method. It can be useful for users to be aware that it is not the foundation model itself that looks up in the database, but a separate search function. It does not help if the foundation model is smart if the search engine is stupid. The model only gets access to the information the search engine pastes.

4.4.4 Open or closed models

Both fine training, few-shot prompting and data access can be used on both open and closed models. However, you should be very careful about what you choose to protect your own data and maintain control over your own systems and costs.

Open models require more technical expertise and must be operated by the enterprise itself. On the other hand, you have control over the process. Closed models can be trained and operated easily through an API, but then you do not have control over the data processing or access to the model itself. Unless explicitly excluded by the provider, it must be assumed that services with closed models collect all the data that goes through the model. Closed models often have a pay-per-use solution, and quickly become expensive with heavy use.

In the AI Act, the enterprise is legally considered to be the provider if the enterprise includes the model in a product under its own name (see section 1.5).

5 Use of AI

Before the AI model is used, the enterprise should consider the opportunities and risks associated with its use. On that basis, one can assess how it is appropriate to use it and whether it should be used at all.[19]

To ensure useful and responsible use of AI, it is important that:

1. The users have good knowledge of what the AI models can and cannot do. Then there is less chance that users base themselves too much or little on the model or use the model in the wrong way.
2. The way AI is used can be defended in terms of safety, ethics and law. For example, the enterprise should focus on "ethical prompting", where users are trained in how questions and context are presented in a way that reduces the risk of a negative outcome, such as a discriminatory outcome.
3. AI is only used where the benefits of its use, such as increased efficiency, outweigh the downsides, such as the costs to people, society and the environment.
4. The enterprise establishes mechanisms to check that AI is used in the way they want.

By using AI responsibly, society's high degree of trust in business and industry and the public sector can be maintained. See also Appendix 5: Checklist for the development or use of AI.

We can distinguish between internal and external use of AI products. Internal use is, for example, use in hiring processes or as an analysis tool. External use is use in the sale, licensing or fronting of AI products to enterprises and individuals, including user interfaces towards customers such as when external people meet a chatbot.

Below is a list of different sources of risk when using AI and what measures the enterprise can implement to limit the effect of the risk source in question. A source of risk may entail one or more of the Five types of risk: legal risk, ethical risk, value risk, task risk or disruption risk. See also Appendix 6: Risk longlist.

5.1 Sources of risk when using AI

Sources of risk	Description	Effort
Illegal data handling	The model collects data, processes data or generates results when used in an illegal way or that otherwise affects the enterprise's legal risk, e.g., through the use of others' intellectual property, personal data, contrary to the enterprise's contractual obligations, etc.	Ensure routines for data minimisation, data transparency, anonymization, data security, regulatory compliance, training of use, analysis of generated results, etc. Consider loss of reputation, liability, fines and contract risk when using such an AI model.
Unethical data handling	The model collects data, processes data or generates results by use in a way that is legal in relation to applicable regulations and other obligations, but which can still be	Ensure ethical and responsible use of AI with associated good processes, defined roles and overviews, a good knowledge base internally, good compliance with

[19] See Chapter 1 for an overview of how such a risk assessment should be carried out

Sources of risk	Description	Effort
	described as unethical and contrary to the enterprise's or society's values.	regulations, open AI practices towards customers, protection of customer data, supplier control, etc.
Data Leak	The model exhibits unwanted data sharing, typically as a result of the model leaking training data to the user of the model.	• Control access to the model. • Check the result from the model. • Making data available during training.
Involuntary data sharing	The enterprise uses a closed model where the model's provider absorbs the user's data during use, which results in the enterprise sharing data when using the AI model (e.g. open ChatGPT).	• Check the terms of use. • Acquire a version that offers a closed environment for the user's input/data (so-called enterprise version).
Failure to explain	A lack of understanding of the model results in the user not being able to produce legally or ethically required explanations of decisions made by or using the model.	Give users (and potential users) of the model a better understanding of the model's functions and results, how these should be interpreted, put in context and followed by humans, and communicate the importance of being able to explain the decisions that are made. Consider alternative AI model. Checklists for justifications can be assessed.
Weakened competitive position	Potential negative consequences that a business may face if its competitors exploit AI more effectively (e.g., loss of market share, reduced earnings, etc.)	Assess how AI affects the enterprise's competitive situation and establish a strategy to meet the competition (e.g., by increasing expertise and improving products/services to ensure its own competitiveness).
Vulnerabilities related to the technology	Covers potential challenges in the implementation and operation of AI systems, including errors or deficiencies in the systems, security vulnerabilities and technology dependence.	Carefully assess the risks and establish robust risk management strategies to ensure the reliable use of AI in accordance with applicable laws and ethical norms.
Third-party risk	The business gets a bad reputation, risks financial loss or weakened creditworthiness by using AI (even when using it responsibly), either due to errors, accidents or irresponsible	Be open and clear about how the enterprise uses AI, and what guidelines and security mechanisms the enterprise has in place to ensure responsible use. Ensure thorough

Sources of risk	Description	Effort
	use of AI by others (e.g., with an AI supplier).	risk assessments, ensure that the AI system used is robust and well-tested, increase the level of competence, closely monitor market conditions and stress test the model regularly.
Provider	The supplier of the AI solution may create operational disruptions in the enterprise's operations and its IT system. The choice of AI supplier can also have a *lock-in* effect, which ties the business to the chosen supplier. It can affect the enterprise's critical functions and be difficult to change in the event of an incident in the future.	Carefully consider which supplier(s) are chosen, open or closed models, how they fit in with the enterprise's suppliers/systems in general and vulnerabilities. Take into account the supplier terms and conditions and the requirements of the Transparency Act in the assessment. See Customizing Foundation models and Model Selection.
Unintentional AI behavior or use	Implementation of AI solutions leads to discrimination, infringement of intellectual property rights, unwanted contractual obligations, etc.	Establish clear guidelines and processes for ethical AI use, ensure an overview of and good documentable compliance with relevant legislation, keep an overview and establish responsibility for AI use, document routines and compliance, and plan for implementation of the AI Act.
Sustainability	Potential negative impacts AI can have on environmental, social and governance (ESG) aspects of sustainability, including power and resource use, social impacts and governance challenges.	Include sustainability risk in AI strategies and include externalities in accounting and reporting, including climate impact and resource use (e.g., use of computing power), social impact (e.g., discrimination and job loss), and impact on decision-making processes must be included in ESG reporting.
Radical disruption	The enterprise's basis for existence can change or, in the worst case, disappear: • Demand for the enterprise's products or services fails through radical disruption, or • Transformative changes in the market and the context in which the business operates will make it impossible to continue as before.	Include AI into the strategy of the business and consider both the short-term and long-term implications of AI. Include AI on the board's agenda and consider measures to limit the risk of disruption, such as radical changes in the nature of the business, changes in the offering of products and services, systemic changes in infrastructure, focus on

Sources of risk	Description	Effort
		R&D, consolidation and partnerships, etc.

5.2 Internal use

When using an AI model internally, it should be considered whether the model will contribute to improving the enterprise's processes, and what negative effects the use of the model may entail.

Sources of risk	Description	Effort
Overuse and addiction ("Over reliance")	• Too much trust in the AI model in various areas leads to systematically poorer decisions. • Overuse of the model leads to a dependency with applications for which the AI model is less suitable and vulnerability to system failures. • The AI model is used to make decisions without the user having sufficient understanding of why the model makes a given assessment.	• Ensure continuous training for the use of AI and provide users (and potential users) of the model and a better understanding of the model's weaknesses and limitations. • Consider the importance of human oversight and verification of AI-generated results. Diversify your supply chain with alternative systems and processes. Conduct continuous evaluations and risk assessments, monitor technology, ensure good data governance, and ethical and legal compliance.
Under-reliance	• Too low confidence in the AI model leads to little use or poorer use. • Underspending can be due to excessive or unfounded skepticism, lack of understanding, or lack of training or experience.	Provide users (and potential users) of the model (i) with a better understanding of the model's features, capabilities, and how the model can be used; (ii) introduction, training and practice in the use of the model, (iii) basis for assessing misconceptions and limitations related to AI.
Misinter-pretation	• The user may misunderstand why the model makes a given decision and how the model's results should be interpreted. • The user may also misinterpret outputs and react negatively to a flawed or incorrect basis.	Give users (and potential users) of the model a better understanding of the model's functions and deliverables, how these should be interpreted and put in the right context and reviewed by humans.
Negativespillover	Even if the model is used correctly, its use may entail unintentional negative systematic impacts on the business or on	Continuously assess the broader implications of the AI strategy and use of AI models, increase employee

Sources of risk	Description	Effort
effect	society (e.g., job loss, unrest, oppressive use of personal data, reinforcement of discriminatory practices and environmental impacts)	competence, develop policies, controls for potential negative impacts, and take appropriate action.
AI dependency	The business becomes dependent on the AI model to make good decisions. This can result in vulnerabilities in the form of the enterprise's goods or services, or internal processes being disrupted if the AI model were to become unavailable.	Ensure a balanced use of AI, consider alternative AI models and backup solutions, maintain human knowledge, establish robust control and monitoring systems, etc.
Operational problems	The potential for loss due to errors, failures or deficiencies in internal processes, people, or systems, where AI systems are involved (e.g., technical failures, operational interruptions, failing data quality and handling, security breaches, lack of knowledge and poor change management)	Ensure thorough testing and validation of AI systems, continuous monitoring and maintenance, effective data management, robust security measures, and good training and support for employees. Ensure a verifiable AI domain for the business (e.g., in your own protected cloud solution) and have an internal group (task *force*) that can quickly understand problems that have arisen and take action.
System failure	AI can have a negative impact on the enterprise's systems or be used to uncover weaknesses that can lead to data breaches, network effects with a potential for damage across functions/infrastructure, etc.	Businesses and authorities must work together to develop standards and protocols for AI security, monitoring and control of AI systems, as well as contingency plans for handling possible systemic failures.

5.3 External use

When an AI model is offered as part of a service to third parties (market) or to users or citizens in society at large (e.g., public services), particular challenges may arise.

Sources of risk	Description	Effort
Reduced quality of product/ service	AI may reduce the quality of a product or service by making it more generic or having lower quality, with associated product liability and a weakened market position.	Assess the effect of using AI and how this affects the product or service. Prepare an overview of the enterprise's important characteristics and the risk of diminishingthem. Also consider what responsibilities the enterprise has under consumer

Sources of risk	Description	Effort
		legislation and the EU Product Liability Directive, etc.[20]
Misinfor-mation	The enterprise has a responsibility to ensure that correct information about the product or service is provided, which is reinforced by the AI Act.	Consider how the product or service is discussed externally/marketed and what responsibilities the enterprise has under the Marketing Act, among other things. Mark the AI-generated result (*watermark*).
Negative societal impacts	This may be because it is misused, or that the use has unintended consequences, or that it is part of a use that causes systemic problems.	Put in place a system to monitor whether the system fulfills the function as intended and how the risk of misuse can be reduced.
Negative impact on people	When using AI, the product or service may have a negative impact on people, e.g. through fake images or news (deepfakes)	Get an overview of how the product or service may be (mis)used and what risk-reducing measures can be implemented to prevent such use.
Reduced cyber-security	The use of AI models can give rise to new security vulnerabilities. Some examples of this could be: • "Prompt injections": Third-party attacks using specially designed prompts (either directly or indirectly via a web page/file the system reads), resulting in undesirable consequences. • "Model denial of service": Third parties send requests to the AI model that take up so many resources that it prevents others from using the service. • "Supply chain vulnerabilities": AI models (especially language models) lead to new dependencies that are difficult to detect. This can apply to everything from the training data, to the foundation model or plugins. • For more comprehensive overviews of such forms of risk, see OWASP's top 10	• Developers and providers should get an overview of new vulnerabilities and how they can be counteracted. • The National Cyber Security Center (NCSC) has developed guidelines for the safe use of AI in collaboration with Norway and 17 other countries.[22]

[20] The proposal is being considered by the EU
[22] Se Guidelines for secure AI system. Se også OWASPs AI Security and Privacy Guide development.

Sources of risk	Description	Effort
	threats for LLMs and for information security in general.[21]	
Model theft	Theft, or copying of the enterprise's AI model.	Make sure to consider the need for the development of robust security routines around access to the model, an overview of weaknesses that may result in model theft, as well as measures should this occur. Consider comparison and possible coordination with the enterprise's procedures for intellectual property rights and trade secrets
Data theft	Theft of the enterprise's data, or data in or at the basis of the model.	Assess the need for the development of robust security routines around access to data/trade secrets and an overview of weaknesses that may result in data theft, as well as measures should this occur. Consider comparison and possible coordination with the enterprise's routines for personal data.
Downtime	Due to integration with an AI model, a service can become less robust, and this can result in increased downtime, or poorer quality.	Consider mapping the risk of weaknesses as a result of integrations, overview of measures that can prevent or repair such weaknesses should they arise.

The various sources of risk and mitigation measures must be assessed in the context of the actual development or use of AI. The current focus on large language models and generative AI creates a need to identify specific sources of risk for this technology. This can be:

i. Deliberate Misuse: unethical or illegal exploitation of generative AI (e.g., fraud or the use of deepfakes to spread disinformation)

ii. Unintentional Misuse: incomplete or incorrect results (e.g., where the model is hallucinating)

iii. Misrepresentation: the results from generative AI are used and disseminated despite uncertainty or lack of credibility (e.g., videos with uncertainty related to the source and use of AI)

[21] Top 10 Threats for LLMs and Top 10 Information Security.

iv. Accident: the result from generative AI is disseminated without the user being aware of errors or lack of credibility (e.g., dissemination of video that appears to be real, but which later turns out to have been created using AI).[23]

In light of the rapid technological development, the enterprise must ensure a continuous overview of how AI affects the enterprise and to what extent the five different risks materialize. Such an approach could result in responsible use of AI.

[23] Overview from Öykü Isik, Amit Joshi, and Lazaros Goutas: "*4 Types of Gen AI Risk and How to Mitigate Them*" Harvard Business Review (31.5.2024), see reference in Appendix 7.

Appendix 1: Relevant user groups – who is the Standard aimed at?

We believe that the standard will be relevant for a wide range of Norwegian enterprises, both private and public, in their own work to develop and use AI responsibly. We have identified the following functions that may benefit from the Standard:[24]

(i) *Business leaders, board members, and other policymakers*: This group needs to understand the broader implications of implementing AI in their operations, including ethical considerations, business risks, and regulatory compliance.

(ii) *Compliance and legal teams*: These functions must ensure that AI applications comply with relevant laws, standards, and ethical norms.

(iii) *Ethics and AI governance teams*: Bodies within organizations dedicated to enforcing the ethical use of AI.

(iv) *CIOs/Product Managers*: They oversee the development of IT-related systems/products, including AI products/services, from concept to launch. They need to understand how AI features and risks affect product strategy and user experience, and how AI systems affect the overall IT architecture and system risk. This includes cybersecurity and regulations such as the EU's Digital Operational Resilience Act (DORA) and NIS 2 Directive.

(v) *AI developers and engineers*: These are the technical teams that design, build, and maintain AI systems. They need guidelines on how to implement ethical principles in their code and understand the risks associated with the AI systems they develop.

(vi) *Risk management specialists*: They are responsible for identifying, assessing, and mitigating risks associated with AI systems. They search for guidelines on potential AI risks and how to integrate risk assessments into business processes.

(vii) *Data scientists*: People involved in data analysis and building machine learning models. They require an understanding of the implications of the data they use and how to design models that are fair and transparent.

(viii) *Business functions*: They may use AI in their business functions (e.g., finance, HR, sales and marketing teams, etc.), and must be aware of how the technology is used, potential hallucinations and biases (i.e., biases and discriminatory practices), and the extent to which personal data and/or individuals' rights are affected (which requires assessments under the GDPR and the AI Act).

(ix) *End users and customers*: The final recipients of AI-powered products and services. They benefit from understanding how AI works, transparency about the use of AI in the service they use (e.g. when using customer support), AI's limitations and potential risks (e.g. related to personal data or other rights) so that they can use these products responsibly.

(x) *Non-governmental organizations:* Organizations that focus on consumer rights and societal impacts of technology, that seek to understand and influence the development of ethical AI policies, as well as the use of technology to achieve its purpose more effectively (e.g., effective allocation of funds to the organization's target audience).

[24] This list is not intended to be exhaustive

Appendix 2: Key Terms and Definitions

Below is a list of key terms and definitions relevant to understanding this Standard:[25]

Idea	Definition
Artificial Intelligence (AI) & AI based systems: *("AI systems")*	Artificial intelligence, or an AI system, is a machine-based system, which operates with varying degrees of self-determination and can be adapted after implementation. An AI system steers by goals and generates results from the input it receives (e.g., a prompt), in the form of predictions, content, recommendations, or decisions that may affect physical or virtual environments.[26]
General Purpose AI Models:	AI model that demonstrates significant general ability and is able to competently perform a wide range of different tasks.[27]
Large Language Model:	A generative AI model capable of understanding and generating natural language and other types of content to perform a wide range of tasks. LLMs are trained on vast amounts of data, enabling them to recognize complex patterns in existing content and generate new content. An LLM may in some cases be classified as a general AI model, but not necessarily.
Generative AI	A type of AI that, based on large amounts of training data, can generate new content, such as text, images, videos, code and other types of data, based on requests from users.[28]
Output	Result or "output" generated by an AI model based on a request from the user (so-called "prompts").[29]
Prompts / input	Request from the user ("input") to an AI model to generate a result ("output") based on the AI model's training data.
Hallucinations	When a large language model (LLM) creates AI results that are inaccurate or incorrect.[30]
Bias	Incomplete or incorrect/skewed results derived from the original training data or AI algorithms.[31] The bias can lead to a violation of equity and non-discrimination principle rooted in Section 98 of the Constitution and the Equality and Anti-Discrimination Act.[32]
Deep Fake	AI-generated or manipulated image, audio, or video content that resembles existing people, objects, places, devices, or events and would appear to an individual to be authentic or truthful.[33]
Responsible AI	Set of principles that provide guidance on the development, implementation and use of AI to mitigate risk, which in EY's

[25] Not an exhaustive list of relevant terms for AI systems and not alphabetical order.
[26] EU AI Act Article 3 (1)
[27] EU AI Act Article 3 (63)
[28] Se f.eks.: Generative Artificial Intelligence (AI) | Harvard University Information Technology
[29] Based on OECD: Updates to the OECD's definition of an AI system explained - OECD. AI
[30] Based on IBM's article: What Are AI Hallucinations? | IBM
[31] Based on IBM's article: What Is AI Bias? | IBM
[32] See also the Equality and Anti-Discrimination Ombud's AI guide
[33] EU AI Act Article 3 (60)

Idea	Definition
	framework are accountability, fairness, trust, transparency, explainability, data security, legal compliance and sustainability.[34]
Ethical AI	For example: i) Respect for human self-determination, ii) preventing harm, iii) Justice, and iv) Explainability.[35]
Risk	The combination of the likelihood of occurrence of injury and the severity of the injury.[36]
The AI Act	The EU AI Act which achieved political agreement in December 2023 and was finally adopted by the EU Parliament in March 2024 and by the European Commission on 21 May 2024, with the last linguistic version of 19 April 2024.
AI Literacy	Skills, knowledge and understanding that allow suppliers, users and affected persons, taking into account their respective rights and obligations under the AI Act, to make an informed implementation of AI systems, as well as to be aware of the opportunities and risks of AI and the possible harm that it may cause.[37]
Intellectual property rights *("IPR")*	Intellectual property rights can be understood as *creative thought work*, and the right to decide over and take ownership of the value of the work.[38] Intellectual property rights include intellectual property[39] with associated copyright, patents, trademarks, trade secrets, etc.
Enterprise	Legal person, which manufactures or offers goods or services. A distinction is made between private and public enterprises, commercial, non-commercial and non-profit.[40]

[34] Based on EYs *'Responsible AI Framework'*: EY's commitment to ethical and responsible AI principles | EY - Global
[35] Based on EUs Ethics Guidelines for Trustworthy AI from 2019: Ethics guidelines for trustworthy AI - Publications Office of the EU (europa.eu)
[36] See the EU's AI Act Article 3 (2)
[37] EU AI Act Article 3 (56)
[38] See the Norwegian Industrial Property Office with references: What are intellectual property values and rights? - Patentstyret
[39] See the Copyright Act
[40] As defined in Store Norske Leksikon: business – Store norske leksikon (snl.no)

Appendix 3: Strategy checklist

1	Purpose	The enterprise must clearly define what it wants to achieve with AI, and how the development or use of AI is linked to the enterprise's strategy. This can include streamlining or automating tasks, improving customer service, innovating products or services, or creating new business opportunities.
2	Develop, buy or wait	The enterprise must decide whether to build AI expertise internally, buy external services from AI suppliers or wait until AI becomes an even greater off-the-shelf product. This choice will depend on the enterprise's existing expertise, resources and long-term goals.
3	Data strategy	AI relies on data to learn and function effectively. The business must develop a data strategy that ensures access to high-quality data, while also complying with privacy legislation and ethical guidelines.
4	Model selection	If the enterprise wants to use a model that has been developed by others, there are various options: one can license a closed AI model, download an open source AI model, integrate (embedded) AI directly into other hardware or software, or subscribe (SaaS) to AI solutions via a website or application without integrating the solution into the enterprise's cloud service. See Chapter 2.6 for more on the trade-offs in this regard.
5	Ethics and responsibility	It is important to consider the general ethical implications of AI use, including discrimination and privacy. The enterprise should establish guidelines for responsible use of AI and mechanisms for transparency and accountability.
6	Regulatory compliance	The enterprise must consider the regulatory context of the development or use of AI, both in relation to existing technology-neutral legislation and future specific legislation related to artificial intelligence.[41] For public agencies, this point will also entail the question of whether they have the necessary expertise and resources to formulate rules, monitor and enforce the regulations.
7	Competence and training	The business must invest in training and developing employees to build the necessary skills to develop, manage, and maintain AI systems.
8	Integration with existing systems	AI can and possibly should be integrated seamlessly with the enterprise's existing technical infrastructure. This requires strategic planning to ensure compatibility and safeguarding of IT security.
9	Scalability	The business must consider how AI solutions can be scaled to meet future growth and changing needs.
10	Measuring success	The effect of AI initiatives should be measured and the strategy adjusted based on the results.
11	Long-term vision	AI technology is evolving rapidly, and the business should have a long-term vision for how it can adapt to new opportunities and risks that arise.

[41] As an example, the use of AI to price insurance products may be in violation of the Insurance Contracts Act because insurance companies must ensure the "correct price" rather than the highest possible price.

Appendix 4: Overview of international regulations

The EU has adopted the most comprehensive AI regulatory framework. In conjunction with the General Data Protection Regulation (GDPR), the EU seeks to protect citizens' fundamental rights and privacy. See more about the EU's AI Act above under item 2.5.1

The United States, on the other hand, has so far adopted a "Blueprint for AI Bill of Rights" from October 2022. This forms the basis for a future bill for the regulation of AI in the US. It is based on a set of guidelines for responsible AI, based on the following five principles: i) Safe and efficient systems, ii) protection against algorithmic discrimination, iii) privacy, iv) information and explainability, v) human alternatives, assessments and safety nets.

Furthermore, in July 2023, seven leading AI enterprises in the United States made a voluntary commitment to the U.S. government to focus on safe, secure, and open development of AI technology.[42] The US authorities have subsequently (October 2023) issued a presidential order[43] that reflects the commitment of enterprises to safe, secure and trustful development and use of AI, which does not entail any legal requirements for enterprises but forms the basis for the development of guidelines and standards related to AI. However, there is no expectation of the adoption of comprehensive AI regulations in the US anytime soon, which means that some states have started preparing their own AI regulations at the state level, such as Virginia.[44]

Other countries such as the United Kingdom, Canada, China, Japan, Korea, Singapore, etc. have adopted various forms of AI regulations. Based on the regulations, we can identify some clear regulatory trends:

- The regulations and guidelines from authorities are generally in line with the principles for responsible AI developed by the OECD and confirmed by the G20 countries.
- The various jurisdictions have generally taken a risk-based approach to AI regulations. This means that they tailor regulations to the expected AI risks, such as privacy, non-discrimination, transparency and security.
- Due to the different applications of AI, some jurisdictions focus on sector-specific rules in addition to sector-agnostic approaches.
- AI rules are designed in the context of other digital priorities, such as cyber-security, privacy, intellectual property rights – with the EU as the leading player.
- Several jurisdictions are adopting regulatory sandboxes as a tool for the private sector to collaborate with policymakers. This is to develop rules that both meet the goal of responsible AI and at the same time take into account the implications of high-risk innovations associated with AI.

In addition, regulatory authorities should, as far as possible, engage in multi-lateral processes to make AI rules between jurisdictions harmonised and comparable in order to minimise the risks associated with regulatory arbitrage. This is especially important when considering rules that govern the use of a global technology such as AI. As an example, we see significant differences between the EU, the US and China with regard to the use of personal data in the development and use of AI models.

[42] These are Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI, see more here.
[43] The presidential order can be read here.
[44] Which focuses on the use of AI by public authorities, read more here.

Appendix 5: Checklist for the development or use of AI

Relevant questions enterprises should ask themselves when developing or using AI:[45]

1. Does the enterprise have an overview of and information about the technology and its risk sources?
2. Does the enterprise have the right expertise to develop or use the technology?
3. Has the enterprise organised itself appropriately to take into account the development, use and impact that affects the enterprise as a result of AI?
4. What strategy does the enterprise have for utilising AI, and how does this relate to the strategy for the business as a whole?
5. Are responsibilities/roles for ensuring an overview of the development and use of AI distributed throughout the organisation?
6. Does the use of AI affect the enterprise's market position?
7. What is the purpose of using AI; Streamlining internally or increasing sales to customers?
8. Do the assumed benefits of the use/development of AI outweigh the potential risks and any risk-reducing measures that must/should be implemented?
9. Which AI model should be used as a basis for the enterprise's use or further development?
10. Should an open or closed source model be used, and what factors must be taken into account in light of this choice?
11. How should the enterprise distinguish between internal and third-party models to control uncertainty?
12. What regulations/requirements apply to the enterprise's use or development of AI, taking into account industry, type of business, products/services and use of data?
13. How is the enterprise's use/development of AI classified under the AI Act, and what requirements apply to the services used/developed?
14. Has the enterprise reflected on the security implications of the use of AI?
15. Has the enterprise considered where data is stored and retrieved from?
16. Is any personal data going to processed in the training or use of the AI system, and has a DPIA[46] been conducted?
17. What opportunities does a threat actor have access to their AI tools?
18. Who has, and can have, access to the information the AI models process?
19. What control mechanisms are in place to uncover errors/abuse/inadequate/illegal/discriminatory use of AI?
20. What processes are in place to take into account ongoing risks in the development or use of AI (governance) and is this adapted to actual use/development?

[45] Not an exhaustive list
[46] Data Protection Impact Assessment, required to assess the level of risk to individuals' personal data.

(i) What measures have been implemented to limit risk? How to ensure continuous improvement in line with use and technological development to uncover shortcomings or weaknesses in the measures?

(ii) Will AI be able to create a dependency relationship with a single AI provider that affects the business and its long-term supplier risk? What happens in the event of downtime on the AI solution/model?

21. Who ensures appropriate training and understanding among the users?
22. Can we create a culture of ethical prompting/use of AI to limit potential risks of discrimination?
23. Have we taken into account users in the enterprise who have disabilities and how this group can take part in the benefits of using AI, and if so, whether adaptations are necessary?
24. Will the use of AI have a negative impact on the environment/climate goals set by the enterprise?
25. Are there country-specific / culture-specific conditions that should be taken into account when using AI?

Appendix 6: Risk longlist

The following list of risks can be identified when developing or using AI:[47]

1. Lack of or incorrect knowledge about the use or effect of AI
2. Lack of or incorrect communication about the use or effect of AI
3. Missing or incorrect data foundation (the AI system may be trained on missing or incorrect/non-representative data)
4. Lack of understanding of the AI system's results
5. Data leakage (via hacks or data in other ways going astray)
6. Breach of trust (use of AI beyond what the system can answer for and/or its functions/strengths)
7. Misinterpretation of the AI system's generated output
8. System errors and Hallucinations (the AI system may have system errors, which can lead to hallucinations and other incorrectly generated results)
9. Loss of work functions (streamlining can lead to loss of work functions and create unrest among certain work functions/professions and trade unions)
10. Surveillance (AI becomes a surveillance system, which can be difficult for both society and the individual to control)
11. Concentration of power (too much power is concentrated in certain enterprises/individuals)
12. Plagiarism (AI systems can lead to copyright and other intellectual property violations)
13. Discriminatory outcome/bias (AI results can reinforce historical data/biases in generating new results)
14. Discriminatory use (the use of AI is not adapted to people with disabilities who potentially lose efficiency gains that are made available to the rest of the business/society)
15. Objective Function that doesn't overlap with your enterprise's goals
16. Reputational risk (development or use of AI can result in significant reputational damage, e.g., incorrect use of AI results, copyright infringement, use of people's/customers' data, etc.)
17. Ethical risk (e.g. use of AI that violates society's or enterprise's values and/or norms, without necessarily violating applicable legal rules)
18. Legal risk (in the event of a breach of relevant legislation, there is a risk of both administrative sanctions and penalties, as well as civil liability for damages both for the enterprise and for key persons such as the general manager and board members). This may entail a breach of:
    - Data protection legislation
    - Data Ownership / Data Protection
    - Protection of confidential information / trade secrets

[47] Not exhaustive and in random order

- Intellectual property and intellectual property
- Discrimination legislation
- Consumer law
- Marketing legislation/information responsibility and product liability legislation
- Labour law legislation (e.g. how AI is used, training of employees, discrimination in the hiring process, etc.)
- Competition law (e.g. pricing algorithms, market dominance, misuse of datasets, etc.)
- Law of damages
- Financial legislation (e.g. external AI solutions/models in critical functions without notification to the Financial Supervisory Authority of Norway and necessary control routines)
- Security legislation (e.g. cybersecurity, etc.)
- Criminal law
- Future legislation such as the EU AI Act, which is expected to be implemented in Norwegian law, and the EU's proposed AI Liability Directive, which may be adopted and implemented in the future.
- Contract terms

Appendix 7: Resources

We've used the following resources to develop this standard:

Norwegian sources:

- Norway's position paper to the European Commission regarding the AI Act
- The Norwegian Digitalisation Agency's guide
- The Equality and Anti-Discrimination Ombud's guide

EU sources:

- Artificial intelligence act | Think Tank | European Parliament (europa.eu)
  - See also: The AI Act Explorer | EU Artificial Intelligence Act
- Liability Rules for Artificial Intelligence - European Commission (europa.eu)
- ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf (europa.eu)
- Policy and investment recommendations for trustworthy AI - Publications Office of the EU (europa.eu)
- Digital Operational Resilience Act (DORA) - European Union (europa.eu)
- The NIS2 Directive: A high common level of cybersecurity in the EU | Think Tank | European Parliament (europa.eu)
- GDPR and the Personal Data Act
- Data Act

Other international sources:

- UNESCO Recommendations for Ethical Artificial Intelligence
- OECD:
  - Principles of artificial intelligence
  - AI System Classification Framework
- NIST AI Risk Management Framework (AI RMF)
- DNV's recommendations for AI systems (DNV-RP-0671)
- Task Force on responsible AI | MiT, Stanford & EY
- OWASP (Security):
  - Top 10 Information Security
  - Top 10 for LLMs
  - AI Security and Privacy Guide
- National cyber security center (NCSC) + 18 land: Guidelines for secure AI system development
- Öykü Isik, Amit Joshi og Lazaros Goutas: 4 Types of Gen AI Risk and How to Mitigate Them (31.5.2024, Harvard Business Review)

Other background literature:

- Inga Strümke: Machines that think - the secrets of algorithms and the road to artificial intelligence (2023, Kagge Publishers)