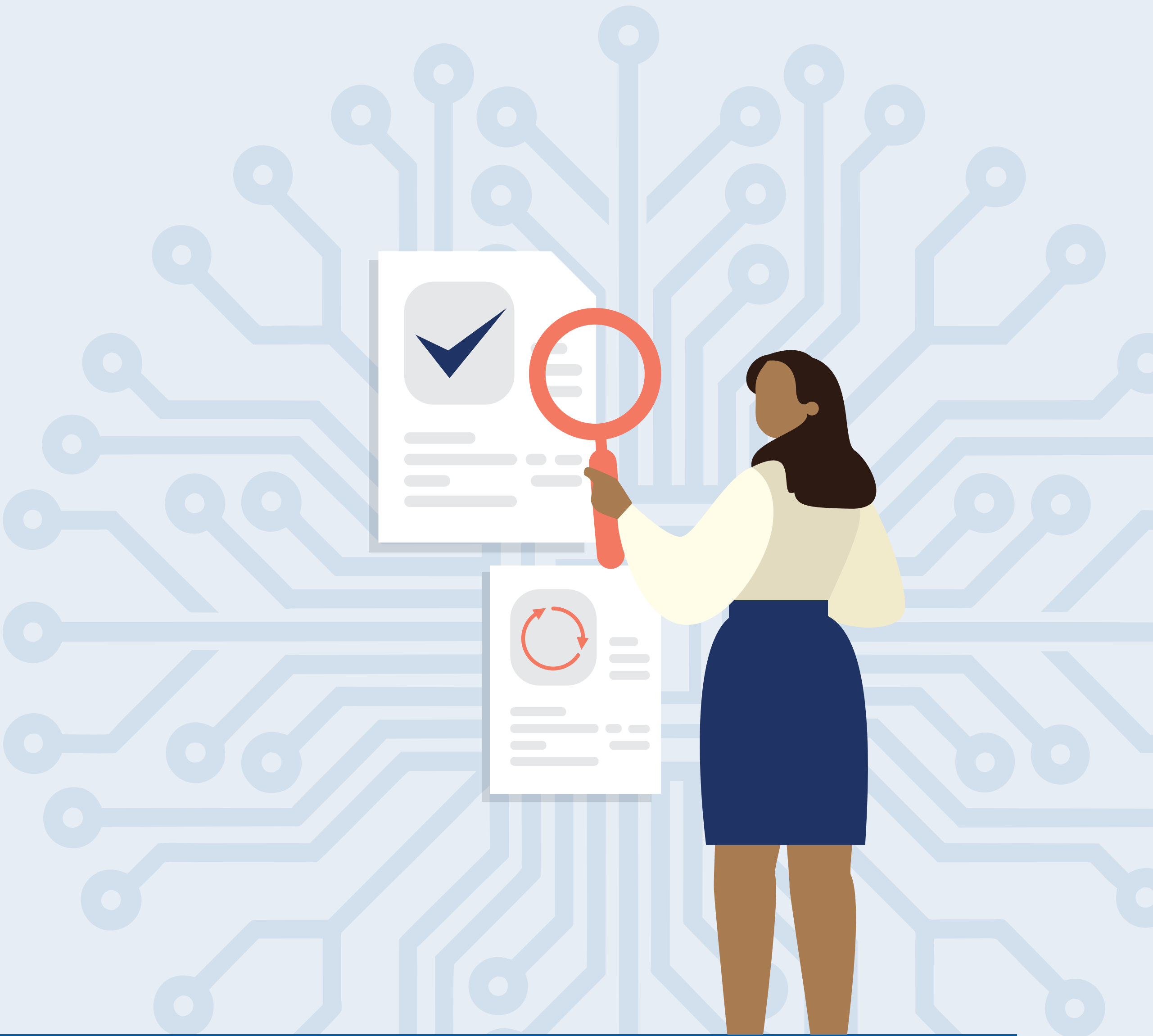
Algorithm Audit



Guidelines for AI Act implementation

Definition of an AI system

December 2024

1. Introduction

The AI Act sets out requirements for the use of artificial intelligence (AI) in the European Union (EU). This product safety legislation ensures the responsible development and use of AI by public and private organizations. This protects the safety, health and fundamental rights of EU citizens. However, implementation of the AI Act raises difficult issues, such as which algorithmic applications fall within the scope of the regulation.

Neither in the public debate surrounding this technology nor within the academic and technical world has a fixed definition been used for the term AI in the past 50 years. Instead, intuitive and unwritten definitions are used. What is considered as AI evolves with what is considered the technological cutting edge: as soon as generally accessible software can perform complex tasks that were previously reserved for 'AI', it is soon no longer considered AI.[1]

However, with the advent of the AI Act, AI is captured in a legally binding definition. The aim of the EU is to use this definition to distinguish AI systems from simpler traditional software systems or programming approaches, thereby providing legal certainty, creating broad acceptance of the technology, and facilitating future-proof practices.[2] The definition adopted by the European legislator is not new: it follows the definition of AI as developed by the Organization for Economic Co-operation and Development (OECD).

Based on this legal definition, organizations must implement the AI Act. However, this turns out to be complicated. Lawyers often have little practical experience with the technologies underlying AI, while engineers are inexperienced with legal definitions. In addition, not all terms in the definition of an AI system carry the same weight. In order to implement the Act, it is therefore necessary to build bridges between these different worlds, whereby not only completeness but also pragmatism must be taken into account.

This white paper helps with these issues. We analyze the core elements of the definition of an AI system from both a legal and statistical perspective, thereby exploring the scope of the AI Act. To this end, an analysis is given of the seven characteristics of the definition of an AI system (sec. 2). Special attention is paid to the concept of inference (sec. 3) and autonomy (sec. 4).

[1] Facial recognition and chess computers were long considered the ultimate examples of AI, but now their applications have become integrated into everyday life and are no longer referred to as such. This phenomenon has been described by Pamela McCorduck as "the AI effect".

[2] See recital 12 AI Act.

Box 1

Disclaimers for compliance with the AI Act

This document is an interpretation of the legal text of the AI Act by NGO Algorithm Audit. No rights can be derived from this analysis. Readers are referred to additional guidelines for the interpretation of the AI Act, which will be published by the European Commission in the course of 2025.

2. AI system definition

The definition of an AI-system is introduced in Article 3(1) of the AI Act. This definition determines the scope of the act. Only systems that meet this definition are covered by the law. The definition can be found in Box 2.

We analyze and interpret the seven concepts, highlighted in color above. The main source for this interpretation is recital 12 of the preamble to the Act. This recital consists of 13 phrases that are listed in the Appendix. The recitals provide insight into the intentions of the European legislator when drafting the legislative text and thus provide guidance on how the concepts should be interpreted. For the interpretation, both the English and Dutch legislative texts were consulted.

When analyzing these concepts, reference is made to the OECD memorandum[3] (hereinafter: 'OECD memorandum') about the definition of an AI system. During negotiations on the AI Act, this memorandum, including earlier draft versions thereof, was used to reach a definition of an AI system in the final legal text. In this light, recital 12 explicitly states that the EU wishes to *"be closely aligned with the work of international organisations working on AI to ensure legal certainty, facilitate international convergence and wide acceptance"*.

We conclude each analysis of the above concepts with a judgment to what extent it can serve as a clear criterion to distinguish AI systems from regular algorithms.[4] An extensive analysis of the concept of inference and autonomy is discussed in 3. Inference en 4. Autonomy

2.1 Interpretation of the definition of an AI system based on recital 12

Recital 12 contains several phrases that help to contextualize the interpretation of the definition of an AI system:

i) *"the definition should be based on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches"*;
ii) *"[the AI system definition] should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations."* – See recital 12 sentence 2.

Phrase i) provides the lens through which we interpret the definition of an AI system: the characteristics in the definition must enable the distinction between AI systems and other software systems. The sentence also functions as a lower threshold with which the legislator indicates that the scope of the definition of an AI system does not concern all programming approaches. 'Simpler traditional software systems' could be understood as simple data processing

[3] Explanatory Memorandum on the Updated OECD definition of an AI system (2024) https://www.oecd-ilibrary.org/science-and-technology/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en

[4] 'Algorithm' as defined by the Netherlands Court of Audit (2021): 'A set of rules and instructions that a computer automatically follows when making calculations to solve a problem or answer a question'.

Box 2

Article 3(1) of the AI Act defines an AI system as follows:

"a machine-based system that is designed to operate with varying levels of autonomy and 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."

in Excel or SQL. However, these programming approaches can also include more advanced data processing that may indeed concern an AI system.

Phrase ii) refers to rule-based algorithms where the rules are set by natural persons. An example of a rule is if age <65 years, then there is no right to senior discount. If the variable age and the threshold 65 years are set solely by natural persons to perform the automatic actions of determining a discount, the rule-based algorithm is not an AI system. This is also the case when this algorithm is used for impactful purposes, such as risk profiling. Phrase ii) has a strong ability to distinguish AI systems from algorithms.

2.2 Machine-based system

Recital 12 states that "*a machine-based system*" from the AI system definition has the following meaning:

"The term 'machine-based' refers to the fact that AI systems run on machines." – see recital 12 sentence 7.

Since virtually all modern software systems or programming approaches use a machine, be it a computer, server or virtual machine (VM)[5], almost all software systems and algorithms meet this requirement.

We therefore conclude that the 'machine-based system' requirement has no distinctive power to separate AI systems from other algorithms, since all modern software systems or programming approaches are machine-based.

2.3 Varying levels of autonomy

Recital 12 states that "*varying levels of autonomy*" from the AI system definition has the following meaning:

"AI systems are designed to operate with varying levels of autonomy, meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention." – see recital 12 sentence 12.

So, there has to be some degree of autonomy. That's why we see autonomy as a factor that can distinguish AI systems from algorithms. In 4. Autonomy the meaning and interpretation of autonomy is discussed in more detail.

2.4 May exhibit adaptiveness

Recital 12 states that "*may exhibit adaptiveness*" from the AI system definition has the following meaning:

"The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use." – see recital 12 sentence 12.

The use of the verbs *may* and *could* indicates that adaptability of an AI system is not a requirement. The OECD also sees adaptability after deployment as optional. In the memorandum, it explicitly states that a system that has been learned once from data is an AI system.[6] Many AI systems that are currently used do not exhibit adaptability after deployment. Facial recognition systems, which the AI Act refers to on several places, are an example where model parameters are not updated in-use but only prior to a software release. In short, AI systems that do not exhibit adaptability in-use can still be an AI system, if the other conditions are met.

[5] A VM refers to a microprocessor that executes algorithms on a PC, laptop, or in a cloud environment. See also 3.32 from ISO/IEC 13522-6:1998 Information technology — Coding of multimedia and hypermedia information

[6] Supra note 3

We conclude that 'adaptiveness' is not a requirement for the AI system definition. As such, it is not a distinguishing factor to separate AI systems from other algorithms.

2.5 Explicit or implicit objectives

Recital 12 states that "*for explicit or implicit objectives*" from the AI system definition has the following meaning:

"*The reference to explicit or implicit objectives underscores that AI systems can operate according to explicit defined objectives or to implicit objectives. The objectives of the AI system may be different from the intended purpose of the AI system in a specific context.*" – see recital 12 sentence 8.

An application always pursues a goal, which can be either explicitly or implicitly defined. The reason this element is included in the definition is to express that an explicit goal is not a requirement for an AI system.[7] For example, through reinforcement learning, AI systems can derive objectives themselves, which are not explicitly formulated but are implicitly captured in the AI system. This is also the case with Large Language Models (LLMs), such as ChatGPT and other applications of generative AI.

The 'objective' requirement has no distinguishing power to separate AI systems from other algorithms.

2.6 Infers, from the input it receives, how to generate outputs

Recital 12 states that "*infers, from the input it receives, how to generate outputs*" has the following meaning from the AI system definition:

"*A key characteristic of AI systems is their capability to infer. This capability to infer refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data.*" – see recital 12 sentence 3-4.

Recital 12 explicitly mentions inference as a key characteristic. Furthermore, inference is explained to refer to both how systems create output and to how the system is created. We conclude that inference is the most important element of the definition to distinguish AI systems from other algorithms. In 3. Inference the meaning and interpretation of inference is discussed.

2.7 Predictions, content, recommendations, or decisions

Recital 12 states that "*predictions, content, recommendations, or decisions*" from the AI system definition has the following meaning:

"*… outputs generated by the AI system reflect different functions performed by AI systems and include predictions, content, recommendations or decisions.*" – see recital 12 sentence 10.

This passage is related to inference, the derivation of output from input, an analysis of which follows in 3. Inference. With regard to "*predictions, content, recommendations or decisions*" this refers to different forms of output that are derived:

1. **Predictions:** This includes estimated scores, rankings, probabilities, labels and classifications. This does not necessarily have to be a prediction about the future, as a prediction can also relate to a previously unobserved data point. The statistical concept of 'estimator' is also called a prediction in this case.
2. **Content:** This includes generated text, images and speech, for example created using generative AI.

[7] Supra note 3

3. **Recommendations:** This includes recommendation systems, such as personalized timelines on social media platforms, search engine results and online advertisements. This category also consists of recommended actions, such as a recommendation for additional monitoring following an assigned risk score for unlawful use of a social welfare services, or a car that recommends changing gear.[8] Scores or classifications to which a fixed policy action or procedure is linked can also be seen as recommendations. Consider: an assigned risk score in transaction monitoring within banks, on the basis of which a work instruction prescribes that additional review must be performed.
4. **Decisions:** This seems to include decisions in the broadest sense of the word, such as the decision to perform an action or procedure, for example a car that brakes automatically for a pedestrian[9], the choice to carry out an investigation, establish someone's identity (verification) or a formal decision as defined in the Public Administration Law.[10] For the public sector, it is important to note that algorithmic output used in the preparatory phase of a decision should also be considered as part of the overall decision-making process and should therefore also comply with principles of good administration, such as the duty of care, duty to give reasons and the principle of fair play.[11] When the output is a recommendation or decision, the concept of 'automated decision-making' from the General Data Protection Regulation (GDPR) is relevant.[12]

The examples (predictions, content, recommendations, or decisions) are an important signal of what the legislator perceives as output of AI systems. Based on this list, a number of types of algorithms can be excluded that do not qualify as AI systems. For example, we establish that algorithms that calculate descriptive (population) statistics, such as averages and standard deviations, are not an AI system. When calculating the average income of a group of natural persons, the output is not a "prediction, content, recommendation or decision". When a statistical model is used to estimate a score for a new data point, then there is a prediction. Following this reasoning, simple data processing and visualization systems do not qualify as an AI system, such as dashboards that display population statistics.

We therefore see characteristics of the output of an AI system as an important factor in distinguishing AI systems from other algorithms, especially in combination with and in relation to the concepts of autonomy and inference.

In order to determine whether an algorithm with a "prediction, content, recommendation or decision" as output is actually an AI system, it is important to investigate how the output is generated. Ways to further investigate the process of the obtained output, in the light of the AI system definition, are explained in 3. Inference.

8 Supra note 3

9 Supra note 3

10 See also Advice on automated decision-making, Dutch Data Protection Authority https://www.autoriteitpersoonsgegevens.nl/documenten/advies-geautomatiseerde-besluitvorming

11 How 'algoprudence' can contribute to responsible use of machine learning algorithms, A. Meuwese, J.Parie, A.Voogt, 2024, Nederlands Juristenblad (NJB) https://algorithmaudit.eu/knowledge-platform/knowledge-base/white_paper_algoprudence/

12 Article 22 GDPR. See also supra note 10.

2.8 Physical or virtual environments

Recital 12 states that *"can influence physical or virtual environments"* from the AI system definition has the following meaning:

"For the purposes of this Regulation, environments should be understood to be the contexts in which the AI systems operate, whereas outputs generated by the AI system reflect different functions performed by AI systems and include predictions, content, recommendations or decisions." – see recital 12 sentence 10.

The physical and virtual environments are complementary. The combination of the two environments is exhaustive. This therefore concerns systems that exert any influence, on any environment whatsoever. This only excludes systems that exert no influence at all, for example because they have not yet been put into use. Furthermore, neither recital 12 nor the OECD memorandum provide helpful explanations for the concept of influence. It seems almost impossible to imagine a system that does not exert influence on any environment.

In any case, the requirement of 'influence physical and virtual environment' is not a criterion by which AI systems can be distinguished from algorithms. The concept of influence is also indirectly discussed in the concepts in 3. Inference and4. Autonomy.

3. Inference

Inference is the key element of the definition to distinguish AI systems from regular algorithms. In this section, several passages from recital 12 are analyzed and related to the AI system definition.

Recital 12 states that inference has the following meaning:

"A key characteristic of AI systems is their capability to infer. This capability to infer refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data." – See recital 12 sentence 3-4.

"The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved.." – see recital 12 sentence 5.

"The capacity of an AI system to infer transcends basic data processing by enabling learning, reasoning or modelling." – See recital 12 sentence 6.

The first and last sentences frame the interpretation: inference is an important characteristic by which AI systems can be identified and it is specifically this characteristic that distinguishes AI systems from other data processing by "learning, reasoning or modeling". Note that only one of these three characteristics need to be satisfied: learning, reasoning or modeling.

These three key concepts are used to analyze the above sentences from recital 12.

3.1 Learning and modeling

Recital 12 states that inference relates to:

"a capability of AI systems to derive models or algorithms, or both, from inputs or data." – see recital 12 sentence 4.

When models or algorithms are derived from data, this is called modelling or learning. Examples are learning the weights of a neural network used for speech recognition or a variable selection algorithm used for risk profiling. Different experts use different terms for this, such as learning, modelling, training or fitting. Regardless of the terminology used, it follows from this passage of recital 12 that inference occurs when a model or algorithm is derived from input or data. Additionally, this passage shows that AI systems must have the capability to infer. This implies there must be a degree of automation in the derivation of models or algorithms from data. If a data analysis is first conducted, for example to determine the average age of a population, which serves as input for domain experts who manually create an algorithm, then this is not a situation in which an AI system infers an algorithm from data.

Recital 12 additionally states:
"The techniques that enable inference while building an AI system include machine learning approaches…" – see recital 12 sentence 6.

In machine learning, a model is 'learned' from a dataset, often called training data. In many cases, statistics are used to calculate model parameters that best fit the available dataset. For data scientists, calculating parameters based on input data is best expressed as the .fit() function, as used in scikit-learn and statsmodels Python libraries. Calculating an average, using a simple formula, is an example of a parameter. So is calculating linear regression coefficients, using a more elaborate formula, or the weights of a neural network using a very complex equation.

Machine learning also includes learning the variables and thresholds of a decision tree for regression and classification. This concerns learning a simple decision tree, but also learning groups of decision trees, such as ensemble-based tree learning. Such as, random forest, xgboost, explainable boosting etc. These are all examples of machine learning.

Whether a data-driven application is called machine learning depends on the domain expertise. An econometrician or statistician will probably not call developing a linear model such as a regression equation or general linear model (GLM) machine learning. However, in this case a model is derived from an available dataset. Based on the text of recital 12, we see no distinction between which technique is used. We conclude that all cases when a model is fitted, trained or learned from data fall under inference.

However, simply deriving model parameters or rules from input data, for example learning regression coefficients, does not make a model or algorithm an AI system. Recital 12 states that inference refers to:

a) *"the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments"*;
b) *"a capability of AI systems to derive models or algorithms, or both, from inputs or data"* – see recital 12 sentence 4.

When learning regression coefficients, b) is satisfied – namely: .fit() – but not a). After all, when learning regression coefficients, no predictions are made for new data points. a) concerns the application of the learned model or algorithm to new data. This process is referred to by data scientists as .predict(), as used in scikit-learn and statsmodels Python libraries. This also relates to the output of an AI system specified by the legislator, namely: *"predictions, content, recommendations or decisions"*. Only after applying this .predict() function is output generated that is required according to the definition. In the case of recommendations and decisions, a score is often first predicted, based on a learned model, after which a recommendation is made or a decision is taken based on this score. A model – based on statistics or machine learning – is an AI system if model parameters or rules are calculated and then a prediction or similar follows. See also 2.7 Predictions, content, recommendations, or decisions, how to generate outputs and 2.8 Physical or virtual environments.

The 'output generation' aspect is an important factor in distinguishing AI systems from algorithms.

3.2 Reasoning: logic and knowledge-based approaches

Inference can also refer to the capacity of an AI system to reason – see recital 12 sentence 6. This indicates that there is a type of system that does not involve learning or modelling, but does involve inference.

This raises the question: which types of algorithms involve reasoning? Recital 12 provides examples of systems that do not fall under this category: *"systems that rely on rules established solely by natural persons to perform actions automatically"* and *"basic data processing"* – see recital 12 sentence 2 and 6.

Furthermore, recital 12 offers little additional clarification regarding the concept of "reasoning." Although, recital 12 mentions the following:

"The techniques that enable inference in building an AI system include … logic- and knowledge-based approaches that infer from encoded knowledge or from a symbolic representation of the task to be solved." – see recital 12 sentence 5.

In logic- and knowledge-based approaches to AI, there is no machine learning; instead, inference is present because reasoning is involved.

Logic- and knowledge-based approaches to AI are also referred to in academia as symbolic AI, as noted in the OECD memorandum.[13] Symbolic AI has been used since the 1980s and 1990s in applications such as chess computers and medical decision support systems. However, with the significant advancements in machine learning, deep learning, and generative AI, this form of AI has received increasingly less attention.

Recital 12 does not provide additional information on the definition and interpretation of logic- and knowledge-based approaches to AI systems. However, the original proposal for the AI Act does include further clarification: *"Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems."*[14] These examples align with interpretations of symbolic AI in academia.

[13] Supra note 3

[14] See Annex I of the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:52021PC0206

To differentiate logic- and knowledge-based AI systems from algorithms, we must identify what sets these techniques apart from *"systems that rely on rules established solely by natural persons to perform actions automatically"* and *"basic data processing"*. We define logic- and knowledge-based approaches using two academic standard works in AI: Artificial Intelligence by Russell and Norvig, and Artificial Intelligence by Poole and Mackworth.[15] In summary, logic- and knowledge-based approaches to AI consist of:

i) **Knowledge base:** An explicit representation of (domain) knowledge. This is often implemented using logic, where knowledge is expressed in propositions and connectives, such as ¬A, A∧B, A∨B, with a proposition (e.g., A) being either *true* or *false*. Other well-known forms of knowledge bases include knowledge graphs.

ii) **Reasoning component:** This component defines how the system can reason about the knowledge in the knowledge base and input data, for instance, using formal logic. This component is also referred to as an *inference engine*. Through the reasoning component, new knowledge and new rules can be derived.

Both components are carefully constructed and require extensive domain knowledge. These approaches are often used when there is a large amount of fixed knowledge and rules in a domain, which can then be reasoned about. For example, in a medical decision support system, the knowledge base may contain medical facts about symptoms, diagnoses, and possible treatments, and the reasoning system can then propose a potential treatment based on input data from symptoms.

Logic- and knowledge-based approaches to AI currently represent a minority. These techniques are mostly used in combination with forms of machine learning. In that case, the system would be considered an AI system due to the use of machine learning, as discussed in 3.1 Learning and modeling. Developers using this type of technology are likely aware that they are employing this form of AI system. We only see the 'logic and knowledge-based approaches' requirement in those rare cases where no machine learning is used, as a key requirement to distinguish AI systems from algorithms.

3.2.1 Reasoning, encoded knowledge and rule-based systems

The AI Act does not provide further clarification of the concept of reasoning beyond the context of logic- and knowledge-based approaches. In the context of the AI Act, reasoning is therefore only associated with logic- and knowledge-based approaches.

It can be argued that in the case of simple manually created rule-based algorithms, reasoning is involved. However, this is incompatible with the explanation provided for the definition of an AI system: *"the definition should be based on the key features of AI systems that distinguish them from simpler traditional software systems or programming approaches"*. If rule-based algorithms are considered to reason, then all types of software systems would reason, which contradicts the intent of the previous sentence. Regardless of whether reasoning is involved, *"rules that are exclusively established by natural persons to automatically perform actions"* do not constitute an AI system – recital 12 sentence 2.

[15] Artificial Intelligence: foundations of computational agents. Poole, D.L. and Mackworth, A.K., 2010. Cambridge University Press. Artificial intelligence: a modern approach. Russell, Stuart J., and Peter Norvig. Pearson, 2016. For an understandable explanation, see also: https://en.wikipedia.org/wiki/Knowledge-based_systems

The passage about "encoded knowledge" – recital 12 sentence 6 – must also be viewed in the context of logic- and knowledge-based approaches. In this context, encoded knowledge relates to the way knowledge is encoded in a knowledge base, as described in 3.2 Reasoning: logic and knowledge-based approaches. Rule-based algorithms, in which human knowledge is encoded, are not practically implemented through a knowledge base (also referred to as a 'knowledge-based approach'). Therefore, the passage on "encoded knowledge" does not apply to the rule-based algorithms we encounter in practice.

4. Autonomy

Recital 12 states that "*varying levels of autonomy*" from the AI system definition has the following meaning:

"*AI systems are designed to function autonomously to varying degrees, meaning that they possess some degree of independence from human involvement and can function without human intervention.*" – recital 12 sentence 11.

To meet the 'autonomy' requirement, there must be *some degree* of autonomy, as also discussed in 2.3 Varying levels of autonomy.

'Some degree' is a minimal requirement: a system does not need to be fully autonomous to meet this criterion. However, the AI Act does not provide further clarification on the concept of autonomy or the different degrees thereof.

The OECD memorandum states that "*the autonomy of an AI system relates to the extent to which a system can learn or act without human involvement*". This implies that any learning algorithm possesses a certain degree of autonomy. In other words, if the inference requirement is met, the autonomy requirement is also fulfilled. Furthermore, the OECD memorandum links autonomy to the different types of generated output, with decisions being the most autonomous and predictions the least autonomous. From this formulation, we infer that the OECD also considers predictions to exhibit 'some degree' of autonomy. By examining the type of output generated by an algorithm (2.7 Predictions, content, recommendations, or decisions) and the type of inference involved (3. Inference), the autonomy requirement can also be examined.

Overall, we conclude that the 'autonomy' requirement does not provide any additional distinguishing ability compared to the other requirements for separating AI systems from algorithms.

Appendix

Recital 12 from the preamble of the AI Act.

.

Sentence 1 – analyzed in 1. Introduction

The concept of "AI-system" in this act should be clearly defined and closely aligned with the work of international organizations dealing with AI, in order to ensure legal certainty, facilitate international convergence and broad acceptance, while providing the necessary flexibility to respond to rapid technological developments in this field.

Sentence 2 – analyzed in 2.1 Interpretation of the definition of an AI system based on recital 12

Furthermore, the definition should be based on the key features of AI systems that distinguish them from simpler traditional software systems or programming approaches, and should not cover systems that rely on rules established solely by natural persons to perform actions automatically.

Sentence 3-4 – analyzed in 3.1 Learning and modeling

A key characteristic of AI systems is their inference capability. Inference capability refers to the process of obtaining outputs, such as predictions, content, recommendations, or decisions, that can influence physical and virtual environments, and to the ability of AI systems to derive models or algorithms, or both, from input or data.

Sentence 5-6 – analyzed in3.2 Reasoning: logic and knowledge-based approaches

The techniques that enable inference in building an AI system include machine learning approaches that learn from data how to achieve certain objectives, as well as logic- and knowledge-based approaches that infer from encoded knowledge or from a symbolic representation of the task to be solved. The inference capabilities of an AI system go beyond basic data processing by enabling learning, reasoning, or modeling.

Sentence 7 – analyzed in 2.2 Machine-based system

The term "machine-based" refers to the fact that AI systems run on machines.

Sentence 8-9 – analyzed in 2.5 Explicit or implicit objectives

The reference to explicit or implicit goals underscores that AI systems can function according to explicit, defined goals, or according to implicit goals. The goals of an AI system may differ from the intended purpose of the AI system in a specific context.

Sentence 10 – analyzed in 2.7 Predictions, content, recommendations, or decisions and 2.8 Physical or virtual environments

For the purposes of this act, environments should be understood as the contexts in which the AI-systems operate, while the output generated by the AI-system is a manifestation of the various functionalities of AI-systems and may take the form of predictions, content, recommendations or decisions.

Sentence 11 – analyzed in 2.3 Varying levels of autonomy and 4. Autonomy

AI systems are designed to function autonomously to varying degrees, meaning that they possess some degree of independence from human involvement and can function without human intervention.

Sentence 12 – analyzed in 2.4 May exhibit adaptiveness

The adaptability that an AI system can exhibit after deployment refers to self-learning capabilities, which allow the system to change during use.

Sentence 13 – not analyzed, because of no specific added value

AI systems can be used on a standalone basis or as a component of a product, regardless of whether the system is physically integrated into the product (embedded) or serves the functionality of the product without being integrated into it (not embedded).

Gemeente Amsterdam

This white paper has been developed in collaboration with Gemeente Amsterdam.

About Algorithm Audit

Algorithm Audit is a European knowledge platform for AI bias testing and normative AI standards. The goals of the NGO are three-fold:

Normative advice commissions	Forming diverse, independent normative advice commissions that advise on ethical issues emerging in real world use cases, resulting over time in algoprudence
Technical tools	Implementing and testing technical tools for bias detection and mitigation, e.g, bias detection tool, synthetic data generation
Knowledge platform	Bringing together experts and knowledge to foster the collective learning process on the responsible use of algorithms, see for instance our AI Policy Observatory and position papers

Structural partners of Algorithm Audit

SIDN Fund

The SIDN Fund stands for a strong internet for all. The Fund invests in bold projects with added societal value that contribute to a strong internet, strong internet users, or that focus on the internet's significance for public values and society.

European AI&Society Fund

The European AI&Society Fund supports organisations from entire Europe that shape human and society centered AI policy. The Fund is a collaboration of 14 European and American philantropic organisations.

Dutch Ministy of the Interior and Kingdom Relations

The Dutch Ministry of the Interior is committed to a solid democratic constitutional state, supported by decisive public management. The ministry promotes modern and tech-savvy digital public administrations and govermental organization that citizens can trust.

Building *AI auditing* capacity

from a *not-for-profit* perspective

www.algorithmaudit.eu

www.github.com/NGO-Algorithm-Audit

info@algorithmaudit.eu

Algorithm Audit

Stichting Algorithm Audit is registered as a non-profit organisation at the Dutch Chambre of Commerce under license number 83979212