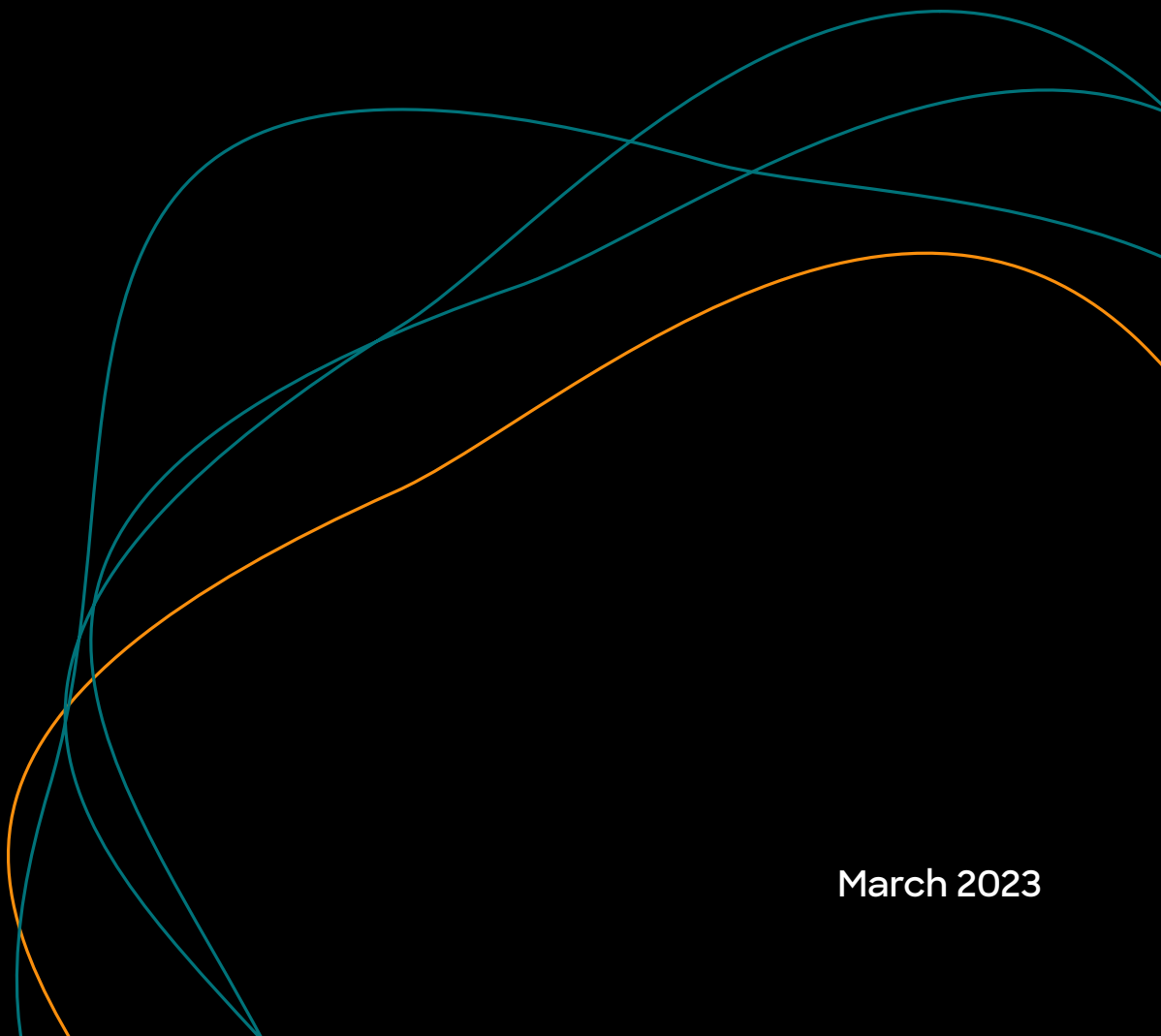
initiative for applied artificial intelligence

AI Act: Risk Classification of AI Systems from a Practical Perspective

A study to identify uncertainties of AI users based on the risk classification of more than 100 AI systems in enterprise functions



March 2023

Contents

Executive Summary

Artificial intelligence is increasingly becoming a part of our everyday lives, whether at home, in industry or in the public sector. The technology poses risks, but also opens up new opportunities. This presents the institutions in Brussels with the challenge of finding a balance between innovation and regulation for AI in the EU.

The upcoming AI Act focuses on the prevention of harm to health, safety and fundamental human rights. Specifically, it takes a risk-based approach, assigning AI systems to a risk class and requiring high-risk systems to meet stricter requirements than AI systems in a low risk class.

Based on the drafts of the EU Commission, the EU Parliament and the EU Council, this study examines the influence of the criteria for risk classification of the AI Regulation on AI innovations in companies and which questions need to be clarified in order to provide more clarity and certainty for planning. In doing so, the study explicitly focuses on the interpretation of the criteria from a practical perspective. At the time of publication of this study (March 2023), the negotiations in Brussels have not yet been concluded, and we hope that our suggestions for a precise classification will be taken up by the negotiators (see next page).

The study at a glance:

- Risk classification of more than 100 AI systems from different enterprise functions such as marketing, production, purchasing, etc, according to the Commission's draft AI Regulation of 2021 and the discussion status in Parliament of early 2022.
- 18% of the AI systems are in the high-risk class, 42% are low-risk, and for 40% it is unclear whether they fall into the high-risk class or not. Thus, the percentage of high-risk systems in this sample ranges from 18% to 58%. One of the AI systems may be prohibited.
- Most high-risk systems are expected to be in human resources, customer service, accounting and finance, and legal. Therefore, fewer companies tend to benefit from AI in these areas.
- Unclear risk classifications slow down investment and innovation. The areas for unclear risk classifications are mainly Critical Infrastructure, Employment, Law Enforcement and Product Safety (Annex II).
- Examining the causes of uncertainty results in concrete recommendations to policymakers and companies to promote responsible AI innovation.

Main causes of unclear risk classifications

Critical infrastructure	• It is unclear whether the European or national definitions “Critical Infrastructure” applies. • It is unclear which asset types and thresholds are applicable for determining Critical Infrastructure, or if they are appropriate for the scalable nature of AI. • It is unclear what should be considered “security components” in Critical Infrastructure, such as in distributed systems like power grids or rail systems.
Employment	• It is unclear how the term “task” is defined in task allocation and how it is distinguished from recommendations, for example. • It is unclear how contractual employment relationships must be structured in order for an AI system to be classified or not classified as “high risk” in this context.
Law enforcement	• It is unclear under what circumstances an AI system is used “on behalf of a public authority” for law enforcement purposes, particularly when companies are required by law to take certain actions, such as in the areas of money laundering, fraud detection, tax returns. • It is unclear which definition of “criminal offense” applies, the national one or a European one. • It is unclear under what conditions a document (or other information) is to be classified as evidence or fact. For example, is it necessary that a legal procedure is already underway, or does it also refer to information that may become evidence?
Safety component (esp. Annex II)	• It is unclear which system boundary applies when determining whether an AI system is used as a safety component? In predictive maintenance, the AI system is often not part of the product being maintained. • It is unclear which definition of “safety component” is to be applied with regard to sector-specific standards (e.g. automotive, medical devices), national laws (e.g. German BSI Act §2 (13) or directives (e.g. 2014/33/EU on elevators and safety components for elevators, Annex III)). • It is unclear whether an AI system for a safety-critical function is not a safety component if there are redundant measures that “step in” and prevent harm in the event of an AI system failure or error.

Foreword

State Ministry for Digital Affairs

Dear Sir or Madam,

Artificial intelligence has made enormous progress in recent years and has become one of the most significant technologies. From chatbots to machine learning and autonomous systems, AI is leading to disruptive developments and is already having a significant impact on our daily lives. For me, therefore, one thing is of particular importance: we as a society must be able to decide at any time whether, where and how we use AI. AI must not be a black box.

That is why I welcome the AI Act's goal of making Europe the center for trustworthy AI. In doing so, the justified interests of promoting innovation and protecting the individual must be appropriately balanced. So we need regulation that protects while leaving enough room for innovation. Otherwise, we not only risk losing touch and becoming technologically dependent on China and the USA. We would also give up the opportunity to carry our free democratic value system into the digital world in a self-determined manner.

The AI Act must become the innovation engine for Europe. We finally need fair market access opportunities so that our European SMEs and startups can compete on an equal footing with developments in other parts of the world.

This study, the only one of its kind to date, uses specific cases from companies to shed light for the first time on the effects of the AI Act regulations in practice. Unfortunately, some of the results are alarming and clearly show the uncertainty surrounding the draft regulation. For example, the classification for the risk-based approach remains too unclear and thus unnecessarily increases the expenses for the companies. The study not only draws attention to the uncertainties in the draft regulation, but also points out concrete options for changes.

There is still time to take countermeasures! The AI Act can still actually become a competitive advantage for Europe. That is what I am working for.

Judith Gerlach, Member of the German Parliament
Bavarian State Minister for Digital Affairs

appliedAI Initiative

The appliedAI initiative was launched in 2017 with the aim of accelerating the application of AI and thus keeping Europe's industry competitive in the AI era. In doing so, we are convinced that we want and need to focus on European values and high-quality AI systems in the most important disruptive technology of our time.

In this context, the EU's AI regulation is the most important legislative intervention to achieve our goal. On the one hand, this calls for a value-based "trustworthy" use of AI, but on the other hand, in the global context, it means first of all additional efforts and complexity for the European industry and thus a systemic disadvantage in the global race for AI leadership.

In order to do justice to this ambivalence, we are committed to point out the practical difficulties posed by the AI Act in a constructive dialog and to identify possible solutions. In this context, I would also like to explicitly emphasize the very good interaction with the European institutions. At the same time, we are working with our partner companies to support the European industry as comprehensively and purposefully as possible with tools, assistance and expertise. This study, together with the survey on the impact of the AI Act on the European innovation ecosystem published a few months ago, is the start of a series of activities on this important and far-reaching undertaking.

For the first time, the risk class of AI Systems in relevant internal functional areas of companies is being analyzed. These AI Systems represent the majority of possible areas of application in corporate processes and therefore provide a good overview of the impact of the AI Act for millions of companies. The often very specific possible uses in products or industry-specific process steps were not analyzed.

We firmly believe that together we can make the AI Act a success story. However, we must keep firmly focused on our goal: To be able to shape, we must remain technologically and economically capable to act. Our innovation ecosystem is our future. In this context, taking opportunities into account is just as important as taking risks into account.

Andreas Liebl
Managing Director of appliedAI

Motivation

The AI Regulation is coming and the European AI ecosystem is preparing for it. The appliedAI initiative aims to inform the negotiations in Brussels and the subsequent implementation with a practical perspective and collaborates with other European and national partners to this end.

Our activities are guided by the following goals:

- Supporting trustworthy AI to protect society and the economy
- Increasing the competitiveness of “AI made in Europe”
- Acceleration of AI development in Europe compared with the USA and China

The risk-based approach is a key mechanism of the AI Regulation with significant implications for the use of AI, because only high-risk AI systems must meet the essential requirements. For providers and users of high-risk systems, the complexity of developing and using AI increases, and so do the costs, which affects the adaptation of AI in practice. The classification rules therefore influence the trustworthiness of available AI systems on the one hand, and the ability of companies to sustainably develop and use such AI systems on the other.

With this in mind, this study takes an exploratory approach to critically examine the proposed classification rules from a practical perspective.

Guiding Questions:

- Which AI systems can be clearly classified and which cannot?
- Which phrases in the classification rules lead to uncertainty?
- What measures reduce legal uncertainty and speed up implementation of the new requirements?

With this study, we would like to encourage members of the European institutions to review the unclear cases and make an exemplary classification.

- Do different people agree with each other?
- Are there already different assessments within the institutions?
- How can different interpretations be avoided in the best possible way?
- What effect do these assessments have on the number of high-risk applications, i.e. do we end up with more like 18% or in the high double digits?

Study design

Data

The data basis of this study are more than 100 AI systems from a public library, which appliedAI created in the context of a BMWK-funded[1] project. The AI Systems are grouped according to business areas such as marketing, production or human resources and described using the following characteristics:

- General description of the situation
- The business problem to be addressed with AI
- Description of the AI system, including functionality and user scenario
- Links and references with background information

Here is the direct link to the open and free database (after registration):

https://www.appliedai.de/de/ki-kompetenz-kurs

Important: The AI use case examples are from public sources, randomly chosen, and purely for illustrative purposes.

Corporate Function	Number of AI systems in this study
Accounting and finances	**10**
Purchasing	**8**
Research and development	**9**
IT and security	**11**
Customer service	**14**
Logistics and supply chain	**11**
Marketing and sales	**14**
Human Resources	**10**
Production and manufacturing	**9**
Legal	**10**
Total	**106**

1 Federal Ministry for Economic Affairs and Climate Action of Germany

Method

Each of the more than 100 AI systems was classified from a practical perspective. The classification rules were operationalized and translated into a methodology by appliedAI and several partner companies as part of a working group on “AI Regulation & Governance”.

At its core, the method consists of four questions to be answered per AI System:

1. Is the system an AI system as defined by the AI Regulation?
2. If yes, does the AI system fall within the scope of the AI Regulation?
3. If so, is the AI system prohibited in the EU?
4. If no, does the AI system fall into the high-risk class?

Here is the direct link to the method template (without registration):

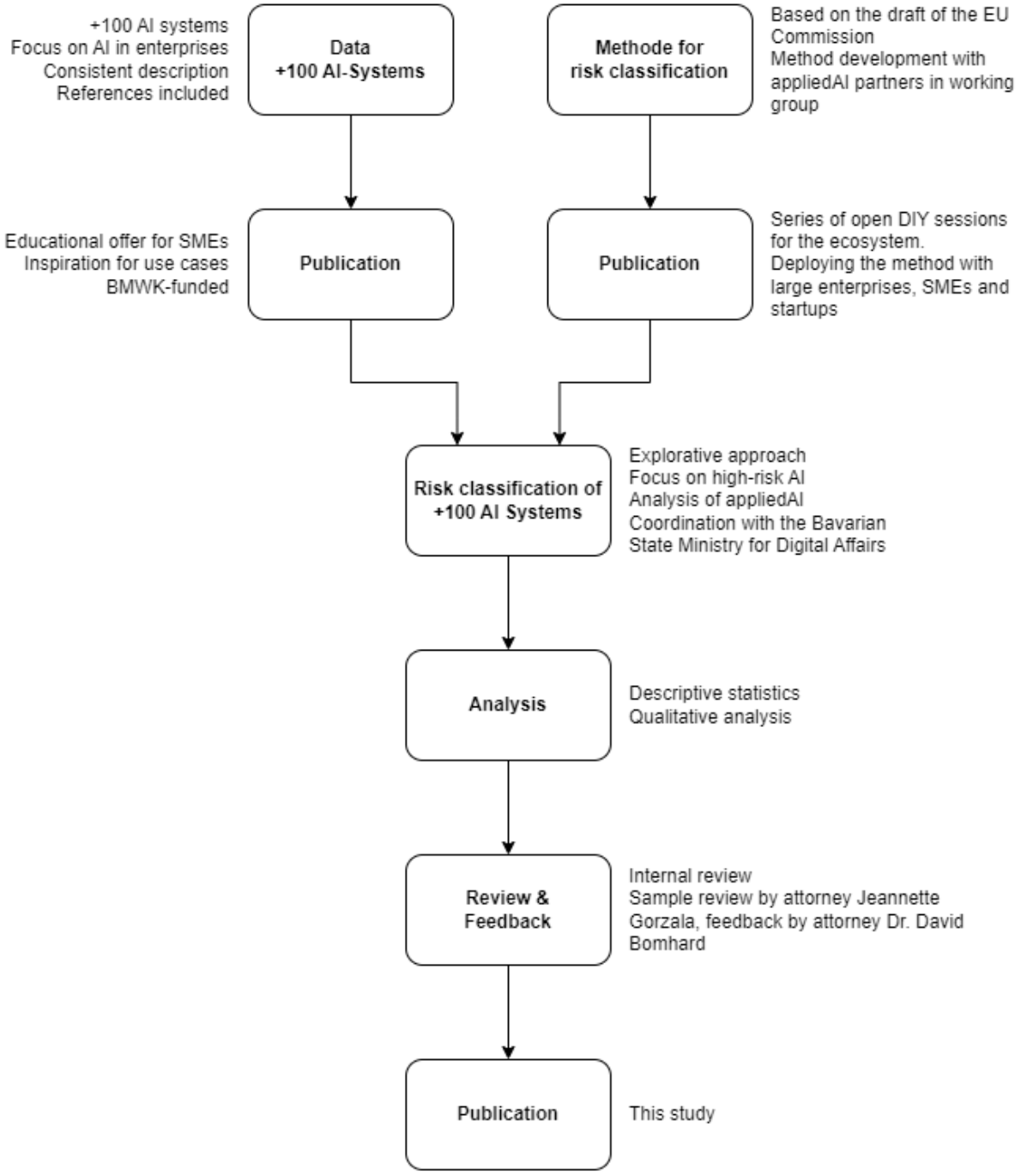
https://miro.com/app/board/uXjVOz16ydQ=/

The classification in this study refers to the proposals of the three institutions involved:

- The initial draft of the EU Commission (April 2021)
- The EU Parliament’s proposed amendments (April 2022)
- The position of the EU Council (November 2022)

Important: The risk classification of AI systems in this study serves the sole purpose of reflecting on the draft AI Regulation from a practical and empirical perspective. It is up to the providers or users of the respective AI systems to determine whether these AI systems are in the scope of the AI Regulation or in which risk class they fall. The risk classification of appliedAI is non-binding and hypothetical with regard to the ongoing negotiations on the AI Regulation.

Study procedure



Risk classification of over 100 AI systems

The risk pyramid

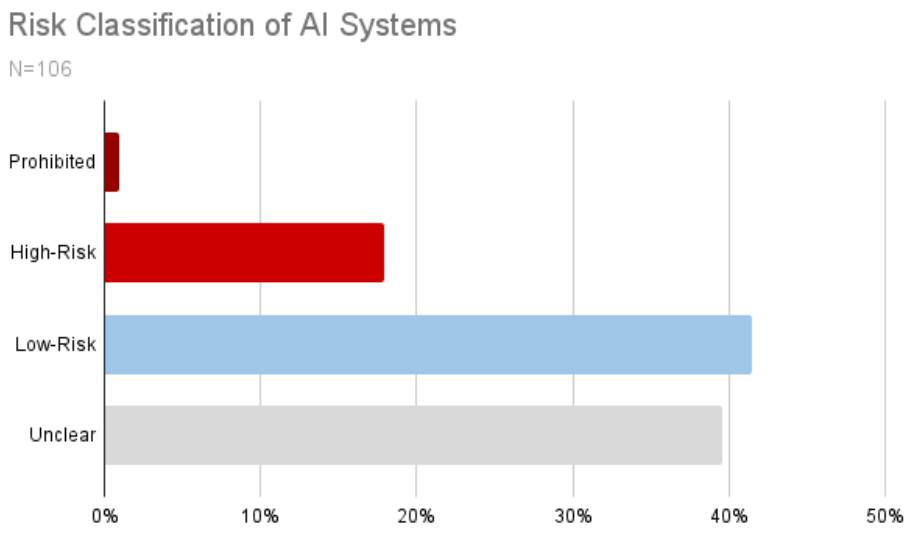
The AI Regulation takes a risk-based approach to keep the legislator's intervention proportional to the risk of an AI system. In principle, the AI Regulation foresees three risk classes:

- Prohibited AI systems (Article 5)
- High-risk AI systems (Article 6)
- Low-risk AI systems not covered by Article 5 or Article 6.

Prohibited AI systems may not be used or made available in the EU (see Title 2 of the AI Regulation). The placing on the market of high-risk AI systems is only possible if the requirements in Title 3 (esp. Chapter 2) are met. Low-risk AI systems are not regulated (a voluntary code of conduct is recommended for them).

Further, Article 52 sets out transparency requirements for AI systems that interact with natural persons, but this requirement applies equally to high- and low-risk AI systems, so they are not listed separately here.

The risk classification of 106 AI systems from the publicly available risk-classification database by the appliedAI Institute for Europe gGmbH[1] yields the following distribution:



The share of high-risk AI systems is 18%, slightly above the maximum value of 15% assumed by the

1 https://appliedaiinitiative.notion.site/Risk-Classification-Database-2b58830bb7f54c9d8c869d37bdb27709

Risk class	Count	Share
Forbidden	**1**	**1 %**
High-risk	**19**	**18 %**
Low-risk	**44**	**42 %**
Unclear	**42**	**40 %**

EU Commission (5-15% was estimated)[1] . This value fits well with the upper end of the assumptions and may be caused by the selection of use cases initially created without knowledge of the AI Act 2020. The additional 40% of unclear cases where an AI system could not be clearly classified as a high- or low-risk AI system brings the potential proportion of high risk AI systems up to 58%. This observation is central to the impact assessment, as most of the requirements of the AI Regulation apply to high-risk AI systems and their providers, for which the overall economic costs and efforts increase accordingly. In this context, we assume that companies are more likely to choose the high-risk category in case of doubt and in order to avoid potential risks. However, in discussions with European institutions, most of the unclear cases could fall into the lower risk category.

Impact of classification rules on AI in enterprises

The AI systems examined in this study are used in general corporate functions, for example marketing, production, finance or human resources. Such AI systems are industry-independent and relevant not only for large companies, but also for small and medium-sized enterprises. Therefore, they have a particularly high potential to create an added value. The large-scale use of AI technology can contribute to significant increases in productivity and thus increase the performance of an economy while using the same or fewer resources. Put simply, “the pie gets bigger.”

The following figure shows the distribution of risk classifications by enterprise function.

1 Impact Assessment, Accompanying the Proposal for a Regulation of the European Parliament and of the Council LAYING DOWN HARMONIZED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, EU Commission, 2021.

Enterprise Function	High-risk	Unclear	Low-risk	Prohibited	Total
Accounting and finances	3	7			10
Purchasing		2	6		8
Research and development		5	4		9
IT and security	2	6	2	1	11
Customer service	4	4	6		14
Logistics and supply chain		6	5		11
Marketing and sales		1	13		14
Human Resources	9	1			10
Production and manufacturing		5	4		9
Legal	1	5	4		10
Total	**19**	**42**	**44**	**1**	**106**

Unsurprisingly, more than 75% of AI systems in human resources are classified in the high-risk category, and more than 25% each in customer service, accounting and finance, and IT and security. Unclear classifications are found in all enterprise functions, but mostly in accounting and finance at over 70%. Only in marketing and sales is the proportion of unclear cases below 25%.

The **risk class of** an AI system affects the likelihood that it will be developed and adapted, or funded. High-risk systems have a lower chance of being implemented because the increase in cost and complexity, due to the additional requirements, raises the barrier to adaptation. A pan-European survey[1] of 113 AI startups and 15 venture capital firms showed that many firms rate the new requirements on data governance and risk management as "difficult" to "very difficult." Further, conducting a conformity assessment is a challenge for many startups. Already now, long before the introduction of the AI regulation, we receive feedback from vendors that companies are very reluctant to apply AI if there is a chance that it might be a high-risk case.

To justify the growing costs, the value proposition of an AI system must be high enough to make the investment in its development worthwhile. Therefore, the added value of the AI system also has an influence on its adaptation in companies. A survey conducted by Bitkom e.V.[2] shows the business areas in which AI is most likely to be used. A total of 539 companies answered the question "In which areas of your company are AI tools used or in which area do you consider future use to be likely?"

Among the companies that were not yet using AI at the time of the survey, areas such as customer service ("for customer retention" with 86%) and IT ("in the IT department" with 82%) are particularly popular. AI is less interesting in the legal and tax departments as well as research and development.

Against this backdrop, AI systems can be expected to be used in particular in enterprise functions

1 AI Act Impact Survey, appliedAI Initiative, 2022, https://www.appliedai.de/hub/ai-act-impact-survey

2 Künstliche Intelligenz – Wo steht die deutsche Wirtschaft?, Bitkom, 2022, https://www.bitkom.org/sites/main/files/2022-09/Charts_Kuenstliche_Intelligenz_130922.pdf

with a low share of high-risk systems and a high potential to add value. In contrast, AI will be found less frequently in enterprise functions with a high risk potential and low added value. Overall, the following trends can be derived for AI in business units (although these are influenced by other organizational factors such as expertise, culture, infrastructure or risk appetite in management):

On this basis, the enterprise functions (points) were assigned to the four quadrants of the matrix. The proportion of high-risk systems (y-axis) comes from the analyses of appliedAI and the interest of companies relates to the responses of those mentioned in the Bitkom study (x-axis). The green dot shows the position of the confirmed[1] high-risk systems (low-risk scenario) for each enterprise function and the gray dot shows the position for the case that all unclear cases are also classified as high-risk systems (high-risk scenario).

Note: The threshold of the quadrants at 50% is hypothetical and it could also be somewhere else. A concrete evaluation of the AI systems per corporate function (e.g., in monetary terms) could lead to a different breakdown.

1 Confirmed in the sense that classification as a high-risk system seems likely.

The **low-risk scenario** describes the case where all unclear cases fall into the low-risk class. Here, several enterprise functions are in the lower right quadrant where the use of AI is attractive, as the majority are low-risk systems with high interest from businesses. This means that investment in development as well as initiatives for application are likely, allowing companies to directly benefit from expected productivity gains. This applies to enterprise functions such as logistics and supply chain, marketing and sales, production, and purchasing. This scenario follows the interpretation of the European institutions involved in drafting the AI regulation.

The **high-risk scenario** describes the case where all unclear cases fall into the high-risk class. Here, the picture shifts significantly due to the high proportion of unclear classifications. As a result, most enterprise functions end up in the quadrant at the top right, which probably leads to a hurdle for investments and adaptations. Only marketing and purchasing would still be in the "attractive quadrant" on the bottom right. This development would put the brakes on the use of AI in enterprise functions such as accounting and finance, customer service, IT and security, or production, because the increased requirements here might not justify the use of high-risk AI. In this case, companies would not or hardly benefit from the potential of AI. This scenario follows the feedback from companies and the general assumption of risk aversity of responsible company divisions.

This **assessment** shows that the large proportion of unclear risk classifications creates a lot of uncertainty in all areas, which can further slow down investment in AI and the already sluggish adaptation of AI in Germany and Europe. A fear of mistakes or penalties in companies matters here too: According to the Bitkom study cited, "violations of data protection regulations" are the second most common concern about the use of AI among the companies surveyed (N=606). This concern may be greater with the AI Regulation than with the GDPR because the AI Regulation is new and therefore more unknown and the penalties for violations are higher. In addition, 49% of the companies surveyed cite "uncertainty due to legal hurdles" as a barrier to the use of AI.

The aim of this study is not that AI systems with a high risk potential should circulate unregulated. The aim is to make a cautious forecast as to what influence the classification rules and, in particular, their interpretation in practice may have on the adaptation of AI in companies and how important precise wording in the AI regulation or legally reliable guidelines will be in this regard.

In order to exploit the potential of AI for companies without compromising the protection of health, safety and fundamental human rights, it is important that the classification rules in the AI Act are clear in order to reduce ambiguities and create planning certainty for investments. Therefore, the following exploratory analysis has the aim of identifying the causes of uncertainties in order to formulate appropriate countermeasures.

Practical review of the classification rules

This chapter examines, from a practical point of view, the extent to which the proposed classification rules can be used to make an unambiguous determination of the risk classes.

For each of the 106 AI systems, we reviewed:

- Is it a high risk system or not?
- Which articles, annexes or recitals are applicable?
- If a clear classification is not possible, what are the causes?

The classification was based on the following provisions of the AI Regulation, which we have summarized in simplified form:

Prohibited AI systems	Article 5 Recitals 7-24 Guiding Questions: 1. Does the AI system use subliminal techniques outside of a person's awareness to significantly influence a person's behavior? 2. Does the AI system exploit the potential vulnerabilities of a particular group to significantly influence their behavior? a. If either 1 or 2 is answered "Yes." Is the behavior change likely to result in physical or psychological harm to this person or another person? 3. Does the AI system evaluate or evaluate the trustworthiness of individuals over time, e.g., based on their social behavior or known or predicted personal or character traits, and does this result in unfavorable treatment that is unrelated or unjustified in context or disproportionate to the behavior or its severity? 4. Does the AI system deploy remote biometric recognition systems in publicly accessible spaces for the purpose of "real-time" law enforcement?

High Risk AI Systems	Article 6 Recitals 30-40 Annex II and III An AI system falls into the high-risk class under Article 6 if (summarized): • The AI system is a product and falls within the scope of one of the regulations in **Annex II** and must undergo a conformity assessment procedure, or is a safety component of a product in Annex II that must undergo a conformity assessment procedure. • The intended use of the AI system falls within one of the application areas of **Annex III**.
Low-risk AI systems	Article 69 The class of low-risk AI systems includes all AI systems that are not prohibited under Article 5 and are not high-risk AI systems under Article 6. The determination of what is a low-risk AI system is not based on a definition or criteria, but on a process of exclusion.

Clear classifications

This chapter shows AI systems that can be clearly assigned or assigned with a high probability to a risk class according to the classification rules of the initial draft of the AI Commission dated April 21, 2021.

The list serves as an illustration and is intended to encourage reflection on whether the classification and the associated requirements are correct or appropriate.

Additionally, this chapter provides a reference point for the ambiguous classifications in the subsequent chapter.

Prohibited AI systems

In the sample of 106 AI applications, we identified one (potentially) prohibited AI system:

ID 93

Enterprise function: IT and Security

Name: Threat detection at major events

Context:

Airports remain a prototypical target for mass-casualty attacks. Many airports across the country and around the world have made major investments in technology, personnel, and processes to screen travelers and other airport passengers. As the second busiest airport in Northern California, passenger traffic at Oakland International Airport (OAK) is on track to surpass the 13.2 million travelers who passed through the airport last year. To accommodate this growth, OAK hired additional staff to keep the airport safe and began exploring innovative solutions in employee inspection methods and equipment. OAK began searching for a new equipment platform capable of detecting a wider range of potential weapons while improving operational efficiency as the number of employees increased. The status quo - a combination of time-consuming, sometimes invasive measures, including walk-through metal detectors and occasional full-body pat-downs - would likely result in long lines at the beginning of each shift and declining morale as a result.

AI system:

An external partner's AI services, trained with a rich set of real-world threat data, continuously distinguish real threats from harmless objects in real time. They become smarter over time as new threat profiles are discovered. They also show security personnel exactly where weapons might be hidden on the person's body or in their pocket, enabling guards to intervene accurately and quickly. The technology uses artificial intelligence and facial recognition software to analyze live footage of approaching airport guests to determine if they are approved individuals, such as regular airport guests, VIPs, employees and others who should be granted access. When a non-permitted person of interest is highlighted, their profile is sent to security officers and a human professional can review and verify the data. The technology claims to allow access to at least one person per second.

Rationale:

The AI system is used to detect people and objects with the potential to pose a threat at a publicly accessible space. It uses biometric recognition and provides information to security personnel, including to manually screen "unauthorized" persons. The exceptions to the prohibition in Article 5(1)(d) (ii) and (iii) may apply*, but there is insufficient information about the AI system to do so.

* (ii) Prevention of threats to critical infrastructure, health, safety and life.
(iii) identification of persons wanted in proceedings.

High Risk AI Systems

Of the 106 AI systems in this study, 19 likely fall into the high-risk class. Here is a selection:

ID 42

Enterprise function: Customer Service

Name: Intelligent Search Example 2

Appendix 3: 1. Biometric identification

Context::

The financial institution partnered with an external vendor in 2017 to provide voice biometrics for authentication in its contact centers. After a smooth experience that resulted in greater personalization and faster resolution for live agent calls, the wealth management company wanted to be even more innovative and further enhance its competitive advantage. They did this by authenticating callers in their Interactive Voice Response system (IVR) even before they were routed to an agent. Their existing voice biometrics solution was extended to include the IVR and the system was tuned to authenticate callers based on minimal voice samples.

Magento is an open source e-commerce platform. Magento's default search is good for basic use cases: it offers auto-completion and synonyms can be added. However, their goal was to provide a better user experience. Magento wanted to closely connect users with strategies, such as marketing search terms or personalizing results, with the right content at the right time.

AI system:

Magento has partnered with an external vendor to provide its clientele with intelligent search on its website. On the site's Support Center page, visitors are shown highlighted parts of the search results that match their query. Through so-called highlighting and snipping, Magento shows searchers why a particular result is delayed, making it easier to find the right solution.

Rationale:

The AI system is a high-risk system under Annex III para 1 lit. a) because it identifies individuals by their voice, which is a form of biometric recognition.

ID 88

Enterprise function: IT and Security

Name: Detection of sophisticated cyber attacks

Annex III: 2. Critical infrastructure

Context:

After a period of corporate restructuring, Energy Saving Trust, an independent organization involved in energy efficiency and clean energy solutions, sought cybersecurity technology to strengthen its overall cyber defense strategy. The Trust was eager to protect its critical assets, including sensitive customer data and intellectual property, from sophisticated and intelligent cyber attacks, and recognized the

need for technology that could detect even the most subtle threats.

In addition, the Trust manages a dynamic and complex network, which naturally makes it vulnerable to potential insider threats. The Trust required complete network visibility to immediately detect unusual behavior, whether from an unsuspecting worker whose system was hacked or a person with malicious intent authorized to use the system.

AI system:

The Energy Saving Trust worked with an outside vendor to develop a platform based on AI technology. The resulting platform models the behavior of each device, user and network to learn specific patterns. The system automatically detects any anomalous behavior and alerts the company in real time. It does this without relying on preset rules or signatures, as most legacy tools do, and therefore is more likely to detect potential threats even if they have not occurred before. Energy Saving Trust can thus detect numerous anomalous activities as they occur and alert the security team to investigate further, while mitigating any risk before real damage is done.

Rationale:

Assuming that the Energy Saving Trust performs tasks in the field of electricity generation and distribution, the AI system is a high-risk application according to Annex III para. 2 lit a) because it serves to protect critical infrastructure in the form of a security component.

Note: According to recital (34) of the Council position of November 25, 2022, this AI system would not be a high-risk application (see "Components intended to be used solely for cybersecurity purposes should not qualify as safety components.").

ID 2

Enterprise function: Human Resources

Name: Analysis of a video interview

Appendix 3: 4. Work and collaboration

Context:

The customer care concept is a central element of the business model of the car rental comparison platform HAPPYCAR. Therefore, it is particularly important to obtain reliable and quick insights into the personality and communication skills of people applying for this functional area. HAPPYCAR was looking for a solution to increase the assessment quality and speed up the recruitment process.

AI System:

HAPPYCAR has integrated Retorio's AI solution for video interview analysis into its hiring process. Based on computer vision and classification techniques, Retorio develops a unique personality profile based on the Big 5 framework and a separate communication profile for each candidate. These two profiles can then be combined into one profile that provides a comprehensive overview of the individual. Individuals applying for a customer

service position at HAPPYCAR must provide a 1-minute video answering the question of why they want to work at HAPPYCAR in addition to their resume. These videos are then analyzed by Retorio's AI solution.

Rationale:

The AI system is used for the evaluation and recruitment of natural persons (applicants) and is therefore a high-risk application according to Annex III, para. 4 lit a).

ID 113

Enterprise function: Accounting and Finance

Name: Risk assessment

Annex III: 5. Access to basic private and public services

Context:

A large financial firm wanted to find early warning signs to identify whether its creditor were likely to become insolvent. The traditional monitoring systems they used screened creditors by checking their bank accounts, remittances, or financial statements. However, when such methods were used, the company was already in financial distress by the time the warning signs were discovered. The financial company worked with Deloitte in the Czech Republic to create an early warning system for credit migrations.

AI System:

Deloitte has developed an AI tool called Eagle Eye that uses open-source intelligence to gather signals from the Internet. The AI software considers any information it finds about the company, customer base or market as a signal. Using machine learning, Eagle Eye then begins to analyze and correlate these signals and can identify specific patterns. AI is able to handle the vast amounts of data on the internet and find correlations between parameters that humans wouldn't even think of. Once these patterns are determined, Eagle Eye constantly monitors the Internet to look for them and provide alerts.

Rationale:

To the extent that the creditors include natural persons and the AI system is to be used for the creditworthiness assessment of these natural persons, the AI system would have to be classified as a high-risk AI system pursuant to Annex III para. 5 lit. b). The exception in Annex III para. 5 lit. b) (or recital 37) does not help here because the AI system is not used by a small or medium-sized enterprise for its own purposes.

ID 107

Enterprise function: Accounting and Finance

Name: Fraud Detection Example 1

Annex III: 6. Prosecution

Context:

Danske Bank has had to pay several billion euros in fines in recent years for not complying with all financial rules and regulations. Although financial crime does occur, most of the cases identified are not fraud, but false alarms caused by outdated IT systems. At the same time, some actual fraud cases are not detected. All suspected cases must be manually reviewed by compliance officers. For this reason, Danske Bank has doubled the number of its compliance staff to 1,700 employees within the last two years. The heavy reliance on manual work increases costs significantly.

AI System:

Hawk:AI combats financial fraud with an anti-money laundering solution based on real-time transaction monitoring that applies machine learning in combination with classic rule-based approaches. Their system analyzes and evaluates large data sets of historical and real-time transactions. Based on insights from historical suspicious cases, the AI system is able to filter relevant cases in real-time and flag them for further investigation by human compliance officers. In addition, Hawk:AI integrates new methods for automatic pattern recognition, enabling the detection of new and unknown types of fraud.

Rationale:

The purpose of the AI system is to assess whether a criminal offense (e.g. money laundering, fraud) has been committed and is thus a high-risk system according to Annex III para. 6 lit a).

Low-risk AI systems

ID 26

Enterprise function: Purchasing

Name: Intelligent supplier selection and management

Context:

In the past, Heidelberger had sourced a rare casting from a single source. Faced with supply shortages and looking for ways to reduce costs, the company needed to find alternative supply companies it could trust and was open to new supply companies from abroad. The process of selecting the right suppliers that are reliable and deliver quality products in a timely manner at the right price can be an arduous task and also requires a lot of analysis and background checks. An additional challenge was selecting a supplier from abroad. Therefore, Heidelberger decided to use Scoutbee's supplier intelligence solution.

AI System:

Scoutbee's Market and Supplier Intelligence solution provides insights into new markets and increases transparency and information quality with its core products. The solution can provide a wide range of potential supplier companies, curate targeted lists, and develop an AI-powered supplier confidence score. Based on this score and Scoutbee's supplier company profiles, companies can qualify and evaluate potential supplier companies faster.

ID 98

Enterprise function: Research and development

Name: Business model review

Context:

A manufacturer of premium vehicles is currently challenged by developments in the fields of electromobility, autonomous driving and car sharing. Expertise in combustion technology, which has guaranteed a certain unique selling proposition for decades, is rapidly losing importance. If autonomous driving and car sharing become established as a trend in the coming years, this could also lead to significantly lower vehicle sales. The manufacturer therefore examined new business ideas from its own employees, customers and external start-ups for feasibility and potential. The challenge was to correctly assess and prioritize the potential of business ideas.

AI System:

The manufacturer sent six employees to a workshop at an external company that offers a business idea analysis solution. After being introduced to the functional principle of the solution, all employees were asked to evaluate their business idea according to their gut feeling. Subsequently, two groups of three people each were formed. Each participant first retreated to a self-assessment of the business idea using the AI tool. The two teams then met again and discussed the results. The AI tool provided concrete starting points to improve the ideas. In addition, in just one day, the team learned the decision parameters of successful venture capital investors and was able to use them to evaluate and improve business ideas.

ID 85

Enterprise function: IT and Security

Name: Data quality monitoring

Context:

Big data sets have big data quality issues. For people dealing with millions of data points, knowing where a change is occurring is a challenge because so many permutations, business metrics, and dimensions exist. These data sets need to be processed on an analytics platform that can efficiently run detection algorithms at multiple steps in the data pipeline to identify data quality issues and changes in business metrics.

AI System:

Anodot's large-scale, real-time AI analytics solution is fully automated at every step of the data collection process (discovery, ranking, and grouping) and provides accurate alerts about changes in key business metrics such as missing data, unexpected data types, zeros where there shouldn't be any, or incorrect records. If these alerts raise suspicions that all is not well with the data, the person responsible can quickly focus directly on the specific problem and consider how to proceed. This multi-member approach can help organizations identify very specific anomalies in data quality, especially those that would be smoothed out or go unnoticed by broader metrics such as averages and company-wide totals.

ID 49

Enterprise function: Customer Service

Name: Cause analysis

Context:

Customer service is often flooded with service tickets. However, when employees have to read every single incoming request, it is often impossible to respond to all issues, leading to an increase in unanswered tickets. A great way to improve customer service is to reduce employee workload on recurring issues. The way to make employees happier is to find the root cause of their problems. Often, support services has a treasure trove of information in the thousands of support tickets they receive each week. It's all in the form of unstructured text data. Since some customers tend to explain their problems in great detail, any root cause analysis based on reading through the tickets and estimates is ineffective and time consuming.

An e-commerce company selling printed products had a large number of complaints about late delivery, and customer dissatisfaction was increasing. Each customer issue had to be handled individually, and it was obvious that these complaints were taking up enormous amounts of time. Since the employees were busy taking care of answering the complaints, the underlying problem could not be identified.

AI System:

With a solution from an external vendor, the e-commerce business used AI-powered root cause

analysis to find interesting correlations and causes that helped it identify a deeper problem beneath the surface. When viewing support tickets, in the context of customer complaints about late deliveries, these tickets were closely linked to their suppliers. If a person complained because of a late delivery, they also named the shipping company. The insights from this can be used to identify shipping companies associated with disproportionate numbers of complaints and take appropriate action.

ID 79

Enterprise function: Logistics and supply chains

Name: Enabling predictive logistics

Context:

Otto is a major player in the German e-commerce market. An analysis of its data showed that returns are less likely if the goods are delivered within two days. In addition, the majority of customers prefer to receive their order at once rather than in multiple shipments. But it's not easy for Otto to cater to these factors, as the company sells products from different brands that it doesn't stock itself. Usually, this means either waiting to ship until all products are collected or sending multiple boxes that arrive at different times.

AI System:

The solution to these problems is to better predict what the customers will buy so that these products can be ordered in advance. To achieve this, Otto uses a deep-learning algorithm originally developed for particle physics experiments at CERN in Geneva. The AI algorithm analyzes around 3 billion past transactions and 200 variables such as past sales, search queries on Otto.de, and external information such as weather forecasts to predict what consumers will buy. The system can now predict with very high accuracy what will be sold in the next month, enabling it to automatically order around 200,000 items each month from third-party brands, without human intervention. At Otto, the AI system has led to a significant reduction in product returns.

ID 68

Enterprise function: Marketing

Name: Brand mention monitoring / social listening

Context:

Somersby, a leading cider brand owned by Danish brewer Carlsberg Group, wanted to optimize and better track their marketing campaigns. They developed numerous hashtag campaigns that successfully engaged fans, and reinforced their success by building strong relationships with bloggers and influencers. For example, when they launched a new Somersby variety in the Polish market, they worked with dozens of bloggers and encouraged people to share content (especially photos) with a special hashtag.

AI System:

Somersby used an AI-powered social listening solution from a third-party vendor to track this campaign and monitor sentiment towards the brand. Thanks to this method, they were able to see that the campaign improved overall brand sentiment and achieved tremendous social media reach. In addition, the new drink became a bestseller in its category.

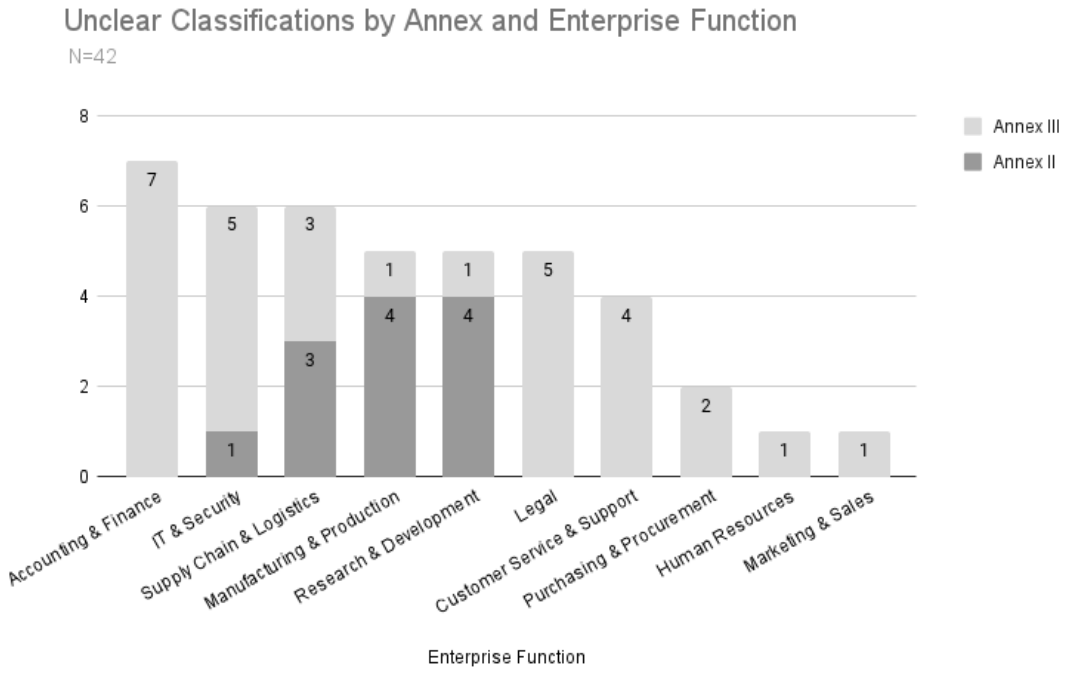
Unclear classifications

This chapter focuses on AI systems for which it is unclear whether or not they should be classified as high-risk systems.

An overview of the 42 unclear cases is followed by an in-depth review of specific AI systems in the areas of Critical Infrastructure, Employment, Law Enforcement, and Annex II. For each application area, the relevant classification rules are first listed. This is followed by the descriptions of specific AI systems, including explanations of why the classification is unclear. The subsequent discussion addresses possible reasons for uncertainty and provides a basis for recommendations for adapting or clarifying the classification rules.

Overview

Of the 42 AI systems with unclear classification, 30 fall under Annex III and 12 under Annex II.



Within Annex III, the most unclear risk classifications are in 6) Law Enforcement, 4) Employment, and 2) Critical Infrastructure. Together with the unclear cases from Annex II, the wording of four parts of the AI Regulation result in more than 80% of the unclear classifications.

The affected AI systems are presented in the following section, including a rationale for the uncertainty. The aim is to gain a more nuanced understanding of the causes of the uncertainties.

Unclear cases according to Annex III:

	Section in Annex III						
Enterprise Function	**2**	**3**	**4**	**5**	**6**	**8**	**Total**
Accounting and finances				2	5		7
Purchasing	1				1		2
Research and development					1		1
IT and security	2				3		5
Customer service			3			1	4
Logistics and supply chain	2		1				3
Marketing and sales		1					1
Human Resources		1					1
Production and manufacturing			1				1
Legal					3	2	5
Total	**5**	**2**	**5**	**2**	**13**	**3**	**30**

Legal basis of the AI Regulation

Classification rules are the basis for distinguishing high-risk systems from other AI systems, so any unclear wording in the AI regulation has a direct contribution to uncertainty among involved stakeholders. Therefore, the following tables contain the relevant classification rules of the areas with many unclear classifications, as a reference for the discussion that follows.

Critical infrastructure
Enterprise function(s): Supply chain and logistics, IT and security
Classification criteria (Annex III; AI Act proposal April 2021) Section 2: Critical infrastructure (+recital 34) (a) AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic and the supply of water, gas, heating and electricity;

Employment

Business Function(s): Human Resources, Customer Service, Supply Chain and Logistics

Classification criteria (Annex III; AI Act proposal April 2021)

3. Education and vocational training (+recital 35):

(a) AI systems intended to be used to determine access, admission or to assign natural persons to educational and vocational training institutions or programs at all levels;

(b) AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions or programs at all levels

4. Employment, workers management and access to self-employment (+recital 35):

(a) AI systems intended to be used for recruitment or selection of natural persons, notably to place targeted job advertisements, to analyze and filter job applications, and to evaluate candidates;

(b) AI [sic] intended to be used to make decisions on promotion and termination of work-related contractual relationships, to allocate tasks based on individual behavior or personal traits or characteristics and to monitor and evaluate performance and behavior of persons in such relationships

Law enforcement

Enterprise Function(s): Accounting and Finance, IT and Security, Legal

Classification criteria (Annex III; AI Act proposal April 2021)

6. Law enforcement (+recital 38):

(a) AI systems intended to be used by law enforcement authorities or on their behalf to assess the risk of a natural person for offending or reoffending or the risk for a natural person to become a potential victims of criminal offences;

(b) AI systems intended to be used by law enforcement authorities or on their behalf as polygraphs and similar tools or to detect the emotional state of a natural person;

(c) AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3);

(d) AI systems intended to be used by law enforcement authorities or on their behalf to evaluate the reliability of evidence in the course of investigation or prosecution of criminal offences;

(e) AI systems intended to be used by law enforcement authorities or on their behalf to predict the occurrence or reoccurrence of an actual or potential criminal offense based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or to assess personality traits and characteristics or past criminal behavior of natural persons or groups;

(f) AI systems intended to be used by law enforcement authorities or on their behalf to profile of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences;

(g) AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.;

8. Administration of justice and democratic processes (+ recital 40):

(a) AI systems intended to be used by a judicial authority or on their behalf to interpret facts or the law to apply the law to a concrete set of facts.

Annex II

Enterprise function(s): Production, supply chain and logistics, IT and security

Classification criteria (Annex II; AI Act proposal April 2021)

Article 6 (+ recitals 30-31):

1. An AI system that is itself a product covered by the Union harmonisation legislation listed in Annex II shall be considered as high risk if it is required to undergo a third party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the above mentioned legislation.

2. An AI system intended to be used as a safety component of a product covered by the legislation referred to in paragraph 1 shall be considered as high risk if it is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to above mentioned legislation. This provision shall apply irrespective of whether the AI system is placed on the market or put into service independently from the product.

Recitals 27-32

Article 3

(14) 'safety component of a product or system' means a component of a product or of a system which fulfills a safety undertaking for that product or system or the failure or malfunction of which endangers the health and safety of persons or property;

Critical infrastructure

Logistics and supply chains

ID 74

Name: Fleet Management Example 1

Context:

Linde is a multinational chemical company with global operations. The company's delivery trucks travel more than one billion kilometers each year. As part of a broader focus on artificial intelligence (AI) and process optimization, the company is now developing solutions to improve the safety of these journeys.

AI system:

Collecting data is a key element of any operational process, as informed decisions cannot be made without analyzing past data. With historical insights, millions of data points analyzed in real time are examined. This results in the prioritization of opportunities and risks, allowing fleet managers and drivers to determine the best course of action in potentially problematic situations. Working with a UK start-up (AI experts in the transport sector), Linde had access to extensive data and began using it to develop a new algorithm. The project focused on external factors rather than information about the drivers themselves. Linde had access to the last 10 years of public transport data, including two million accidents described in police reports, road topology data, weather data, road construction data and traffic data, as well as Linde's own driving records. Machine learning made it possible to identify correlations between different factors, remove irrelevant information and predict what is most likely to happen under certain conditions.

What is unclear?

The training data contains safety-critical data (including police reports) in order to use the AI system to predict possible dangers for truck drivers. It is unclear whether the AI system is considered a safety-critical component in the management of road traffic according to Annex III para. 2 lit a), as a failure or malfunction may lead to an increase in damage.

ID 75

Name: Fleet management example 2

Context:

American multinational technology company Amazon uses many different transportation services to deliver packages. Amazon has long been criticized for pushing its drivers to make up to 200 deliveries a day, which many believe is an unreasonable demand that can lead to tired drivers taking risks. Instead of reducing these intense schedules, the company has begun using AI-equipped cameras to alert drivers when they are breaking traffic laws or engaging in unsafe driving practices.

AI system:

Amazon is installing the Driveri platform from San Diego-based startup Netradyne in its vehicles. Their cameras use four lenses that film the road, the driver(s) and both sides of the delivery truck. The cameras, which are 100 percent operational, do not record audio and cannot be used to observe the driver in real time. They have artificial intelligence that identifies 16 signals based on what is happening around the vehicle and the actions of a driver. Anything illegal, such as failing to stop or driving too fast, triggers audio responses, including "No stop detected" and "Please slow down." Unsafe driving, such as braking too hard, does not generate audio warnings but is captured in footage. This, in turn, is uploaded to a secure portal for Amazon to review. While the cameras don't provide a live feed, some signals can prompt Amazon to contact the driver. For example, if a yawn is registered, the camera will indicate to pull over for 15 minutes. If driver do not do this, presumably because of deliveries that need to be completed, the supervisor could call and ask them to stop for a while.

Note: This application has been significantly criticized and poses some ethical challenges, as some drivers inside see a threat to their privacy.

What is unclear?

One of the purposes of the AI system is to prevent accidents on the road. A failure or error of the system can contribute to an increased danger for drivers. It is unclear whether the AI system is considered a safety-critical component in the management of road traffic in accordance with Annex III para. 2 lit a).

IT and security

ID 87

Name: Detection of sophisticated cyber attacks

Context:

A few years ago, a server breach at commodities trader ED&F MAN Deutschland GmbH was a wake-up call to the increasing success of cyberattacks and the associated risk to sensitive data. An independent assessment led the company to significantly improve its cybersecurity processes and tools and train employees.

AI system:

The company was looking for an AI-based threat detection and response platform from a third-party vendor. This collects and stores network metadata and enriches it with unique security insights. The platform uses this metadata along with machine learning techniques to detect and prioritize attacks in real time. This helped ED&F MAN Holdings detect and block multiple man-in-the-middle attacks and stop a crypto-mining program in Asia. It also found command-and-control malware that had been hiding for several years.

What is unclear?

The AI system supports the protection of a company's digital infrastructure, but it is unclear whether or which companies qualify as critical digital infrastructure within the meaning of Annex III para. 2 of the AI Regulation. The German BSI website[1] lists companies in the special public interest, which includes Germany's largest companies in terms of domestic value added, as well as major supplier companies for these companies.

ID 90

Name: Vulnerability Management

Context:

For business advisory services provider Aprio, risk-based vulnerability prioritization was a time-consuming challenge, requiring staff to manually assess what was important in the unique context of each environment and be able to measure remediation results for each vulnerability. The challenge was exacerbated by the complexity created by hybrid and multi-cloud infrastructures, where organizations

1 https://www.bsi.bund.de/EN/Themen/KRITIS-und-regulierte-Unternehmen/Weitere_regulierte_Unternehmen/UBI/ubi_node.html

management.

AI system:

AI enables threat detection software to think like a hacker. It can help software identify vulnerabilities that cybercriminals would normally exploit and report them to the user. Unlike traditional methods, AI also enables threat detection software to better locate vulnerabilities in user devices before a threat has even occurred. AI-powered security goes beyond traditional methods to better predict what a hacker would consider a vulnerability. Working with an external vendor, Aprio is able to automatically detect assets and apply advanced machine learning to assess the risk that vulnerabilities pose in the context of a given environment. The assessed Health Score also tracks the risk profile of customers and provides a continuous, tangible measure of vulnerability management and remediation efforts rather than a snapshot.

What is unclear?

The AI system supports the protection of digital infrastructure of various companies according to Annex III para. 2 lit a) by detecting vulnerabilities. The companies that make use of this service may also include companies from the critical infrastructure sector. With the same intended use (vulnerability detection), the AI system can be classified as a high-risk or low-risk application depending on the user. It is unclear whether the classification is now up to the provider (e.g., by restricting it to low-risk applications only) or whether the AI regulation stipulates a classification as high-risk applications.

Note: Recital 34 of the draft EU Council would classify the AI system as a low-risk system ("Components intended to be used solely for cybersecurity purposes should not qualify as safety components").

Employment

Continuing education and promotion

ID 8

Enterprise function: Human Resources

Name: Individual learning paths in personnel development

Context:

Although technological advancements have revolutionized L&D (Learning and Development) in organizations over the past decade, there are still some common issues facing L&D professionals, with one of the biggest challenges being the lack of personalized learning. Employees have started to expect something different when they come to work. They want a personalized experience, not a standard experience. They want processes to be tailored and work for them.

Another important factor is that each person has a preferred learning style and learns most effectively using a particular method. This can be through video tutorials, written content, in-person training, gamification, audio-guided presentations, or more.

IBM's internal surveys revealed that several executives were struggling to keep their staff's skills up-to-date and relevant in the face of rapid digital transformation. In addition, they also discovered that there had been an expansion of job roles and that there was a need to address these multidimensional job

roles and changing demographics in the workplace. They were looking for a solution that would work as a multidimensional solution that connected employees, stakeholders, content, services, and provider with a central digital platform to address numerous different roles and needs. Traditional top-down learning management, which decides who needs to know what, offered limited options for employees.

AI system:

Using AI, IBM developed "Your Learning," a rich, personalized digital marketplace for learning. This allowed employees to navigate to the learning content most popular with their team members, sign up for targeted learning channels, and explore the skills and awards they needed to prepare for the company's most attractive roles.

The "Your Learning" platform also helped address workforce demographics, improve the employee learning experience, promote career transparency and improve social compatibility in the organization. A learning chatbot is also available 24/7 to answer questions. As a result, IBM's AI-driven learning platform saw an increase in enrollments and course completions, accelerating strategic skill acquisition.

AI-powered analytics and recommendations can be used to personalize each employee's learning needs. Customized courses can also be developed based on each team member's skills, progress, learning needs, skill requirements, and previous learning successes.

What is unclear?

The AI system generates, based on existing skills and learning outcomes, individual recommendations for further training opportunities. This 1) assesses learning outcomes and 2) is used to guide the further learning process (cf. Annex III para. 3 lit a) "AI systems intended to be used … to assign natural persons to … programs at all levels; to steer the learning process of natural persons"), but it is unclear whether this also applies to internal/non-formal education offerings.

Further, the AI system could fall into the high-risk class because employees' skills are evaluated with an eye toward promotion to a "attractive" position (see Annex III, para. 3 lit b) "AI intended to be used to make decisions on promotion and termination of work-related contractual relationships …)"

With recital 35, the decisive factor should be whether the respective AI system can "decide on the course of a person's education and professional life". The relevance and the effects of the AI decision are probably decisive, but it is unclear what the standard is here.

Task allocation ("to allocate tasks")

ID 44

Enterprise function: Customer Service

Name: Chatbot Example 1

Context:

MAGGI is an international brand of spices, instant soups and pasta. It sells a variety of products around the world. The company goal was to increase customer loyalty. It realized that there is one question that customers are interested in every day: What should I cook today? Therefore, the MAGGI cooking studio wanted to help its customers with recipes and cooking tips.

AI system:

MAGGI has developed a chatbot called KiM ("Kitchen Intelligence by MAGGI") that customers can interact with via Facebook Messenger or WhatsApp. Customers can specify what they are looking for based on ingredients they have at home, dietary preferences and restrictions, difficulty level and preparation time. KiM then sorts out recommendations from 2500 different recipes. KiM can also answer questions about cooking and explain, for example, the best way to peel a pineapple. KiM uses NLP and machine learning to understand logical structures of a conversation, search queries and automate them over time. In the process, KiM learns and stores user preferences, becoming smarter and more helpful from dialog to dialog.

What is unclear?

The AI system gives users individual recommendations according to their preferences (cf. "individual behavior or personal traits or characteristics") and then gives instructions on what to do, e.g. how to peel a pineapple. The AI system appears to be primarily aimed at customers of the spice manufacturer and not at employees (although this is not mutually exclusive), but in view of the wording of Annex III para. 4 lit b), the AI system in question cannot be clearly described as "not high risk" because it is an "AI system intended to be used to … allocate tasks based on individual behavior or personal traits or characteristics …".

ID 34

Enterprise function: Customer Service

Name: Automatic Call Management / Intelligent Call Routing

Context:

Swisscom is a major telecommunications provider in Switzerland. It is the leading provider of mobile, network, Internet and digital TV services for businesses and private customers in Switzerland. Essential to the company are 4,000 sales and customer service employees who handle more than 50 million contacts a year, mostly incoming calls - in German, French, Italian and English. They also handle e-mails, chats and letters.

AI system:

Through AI, Swisscom has better matched customer calls with the best-performing agent:s for different types of interactions. By switching between traditional and predictive routing, the company was able to accurately measure the effect. Average handle time was reduced by 3%. In addition, the company uses intelligent call routing to not only reduce average handle time, but also to ensure that customers are connected directly to agents with the right knowledge and skills. There was no negative impact on other KPIs, such as speed of answer and number of abandoned calls.

What is unclear?

The AI system is a so-called Decision Support System (DSS), which suggests courses of action based on the problem diagnosis. The AI system recommends the next best course of action and routes incoming calls to the most appropriate call center agents. However, it is unclear whether this type of task allocation falls under Annex III para. 4 lit b of the AI Regulation.

ID 76

Enterprise function: Logistics and supply chains

Name: Fleet Management Example 3

Context:

Linde supplies the CO2 cylinders used in pubs for beer dispensers and other beverage dispensers to give them that fizzy goodness. In order to satisfactorily meet customer requirements and provide reliable customer service, Linde had to make additional deliveries to pubs if the pub owner requirements were not met in the first round of deliveries. In addition, in some cases, cylinders were moved around unnecessarily and redundant deliveries were made to pub owners. Analyzing previous data, Linde found that 350,000 gas cylinders per year were being unnecessarily driven around and delivered to different parts of the UK, where Linde was testing the solution.

AI system:

AI algorithms use order history data and combine it with real-time data on other external factors to create a more accurate demand forecast. This improved demand forecast is used to create an optimized delivery schedule that is highly likely to be aligned with customer needs. Such an optimized schedule can help reduce overstocks and understocks for both buyers and suppliers. In addition, algorithms can be tailored to a specific buyer by using that person's order history and combining it with real-time data on other factors such as local events, regional holidays, and relevant weather conditions.

Linde's digitization team used historical data on order information from over 25,000 customers and used AI to determine the influence of other external factors such as weather, local events, holidays, location of pubs and sporting events and their impact on beer consumption in pubs, which then affected the amount of CO2 needed. The digitization team also mentioned that it could have a "customized delivery algorithm" for each customer (in this case, pub owners) that would help deliver the right number of cylinders that a pub owner needs and also deliver them at the right time.

What is unclear?

The results of the AI system influence or determine the routes and travel times of the employees responsible for delivering the CO2 cylinders. The algorithm is also based on individual behaviors and characteristics of natural persons, but those of pub owners and their clientele rather than employees. Therefore, it is unclear whether or not the AI system constitutes a high-risk application in the sense of Annex III para. 4 lit b).

Law enforcement

Prediction of crimes "on behalf of"

ID 106

Enterprise function: Accounting and Finance

Name: Intelligent risk assessment

Context:

Auditing seeks to ensure that the accounts of companies are properly kept, as required by law. Auditors examine the statements before them, obtain evidence and evaluate the statements in their audit

report.

Audit knowledge is to a large extent tacit knowledge that individual professionals have acquired through experience. When formulating the risk strategy, an auditor's knowledge from previous cases is very valuable. Deloitte wanted to make the tacit audit knowledge of individuals more accessible to the entire audit team. To enable knowledge and experience sharing, they began developing the Guided Risk Assessment Personal Assistant AI tool, or GRAPA.

AI system:

GRAPA helps auditors to distinguish a chosen strategy from all other previously used risk strategies. It uses a Deloitte database of 10,000 cases, and each case contains an average of fifty risks. GRAPA is not a standalone application; rather, it is added to the software that accountants use when determining the risk strategy. "It's like asking a second person to read alongside you," explains Van Gool (Audit Innovation Leader, Deloitte). "But the advantage is that this second person has the combined expertise of Deloitte." He emphasizes that the auditor remains accountable for the risk strategy and audit methodology chosen. "GRAPA highlights what has happened in similar cases. But if a company's situation is special or unusual, it is of course up to the auditor to adjust the approach accordingly.

What is unclear?

The AI system is used in the context of an audit to identify potential risks based on similar cases. Companies are obliged to submit a proper tax return and violations can lead to a criminal offense. The AI system helps to avoid such violations. In other words, it generates a prediction as to whether a particular circumstance can lead to a criminal offense.

It is unclear whether the recognition of a possible criminal offense should be understood here as "on behalf" of an law enforcement authority (cf. Annex III para. 6 lit a)), because companies are legally obliged to prepare a tax return.

ID 108

Enterprise function: Accounting and Finance

Name: Fraud Detection Example 2

Context:

AI-based monitoring of transactions in real time can help financial institutions combat money laundering and payment providers detecting fraud. Data generated by real-time payments is fed into the AI system, which then identifies suspicious transactions, stops their processing, and flags the transaction for further review by human compliance officers. The fraud detection system is based on AI algorithms that recognize patterns and identify connections within the data, which are then clustered and classified. Over time, the system becomes accustomed to the data and detection accuracy increases.

AI system:

Worldline, committed to the success and security of its customers, led A.S. Adventure to an innovative solution - Fire by Fraugtser. Fire enables intuitive yet sophisticated fraud detection rule writing, translating human thought processes into unambiguous rules. In addition, Fire allows users to test rules before they are deployed, eliminating uncertainty in rule creation and ensuring accurate performance. The move to Fire enabled A.S. Adventure to easily write and test fraud detection rules. By leveraging Fraugster's AI score, the company was able to reduce false positives to correctly identify good and bad

customers. Running simulations before a rule went live allowed A.S. Adventure to learn how effective a rule could be. Rule performance was improved, eliminating the need for manual reviews and saving valuable risk management time.

What is unclear?

The AI system supports a financial institution in implementing legal requirements to prevent money laundering. It is unclear whether the detection of a possible criminal offense should be understood here as "on behalf" of an law enforcing authority (cf. Annex III para. 6 lit a)), because companies are legally obliged to do so, e.g. by the Anti Money Laundering Act.

ID 109

Enterprise function: Accounting and Finance

Name: Automated expense check

Context:

Electrolux's automation challenge was to increase centralization and improve processes while providing a seamless experience for its many business travelers. The company also needed to maintain its high standards and goals for regulatory compliance. Electrolux manually audited 100% of T&E (Transport & Environment) claims and ensured timely, accurate reimbursements - a thorough but time-consuming and repetitive process. Expense claims were first approved by a manager and then reviewed line-by-line in the SSC. Rejected claims went through the process again, sometimes repeatedly. Receipts appeared in different languages and reports showed varying degrees of compliance with T&E policies. Duplicates were difficult to detect, additional approvals slowed operations, and it was impossible to get an overall picture of repeat offenders. Some auditors did not have the knowledge and experience to find all the errors and anomalies, and too much time was spent on low-risk claims submitted under the guidelines. Electrolux was looking for a solution that would automate the process and allow its auditors to focus only on T&E claims that required a higher level of attention.

AI system:

Electrolux searched for an innovative solution for a long time before choosing AppZen. Expense Audit from AppZen could be integrated with Electrolux's expense automation system to audit every line item in expenses in real time. With its high level of flexibility, AppZen's AI system can provide Electrolux with key information from receipts to identify any major anomalies such as duplicates, out-of-policy expenses or overcharges and comply with required policy rules. The AI system autonomously identifies line items and their expense types and assigns each transaction to the appropriate employee. This improves compliance enforcement and financial regulations.

What is unclear?

The AI system supports the company in complying with financial regulations. A breach of these regulations may lead to a criminal offense. It is unclear whether the detection of a possible criminal offense should be understood here as "on behalf" of an enforcing authority, because companies are legally obliged to do so (cf. Annex III para. 6 lit a)).

Evaluation of documents for legal proceedings

ID 54

Enterprise function: Legal

Name: Intelligent contract management

Context:

To comply with international regulations, companies with leases must go through thousands of contracts one by one. This is an immense task, with an analyst spending around 90 minutes on each contract. In 2019, for example, nearly all leases had to be accounted for under the new IFRS 16 accounting standard. For a telecommunications company that leases every pole and every piece of land on which that pole sits, this meant going through hundreds of thousands of contracts in every possible language.

AI system:

This time can be drastically reduced using machine learning technologies. To help companies with tasks like this, a consulting firm has developed a user-friendly application that can be used by analysts to review contracts. The application has a bot that can be fed a set of contracts. The bot gives the analyst suggestions for dates that are needed, such as the start date of a contract. The analyst sees the highlighted suggestion and indicates whether it is correct or not. The bot learns from this, which results in subsequent contracts being analyzed a little smarter each time and the reliability of its predictions increases.

What is unclear?

The AI system searches contracts for specific content and facts, which are then legally evaluated, e.g. whether a contract needs to be extended in order to comply with a financial law. It is unclear whether, in the case of a lawsuit, such an AI system would be evaluating evidence under Annex III para. 6 lit d) or interpreting facts under Annex III para 8 lit a), and thus would be a high-risk application. The exception in recital 40 (April 2021 draft) for simple administrative tasks ("ancillary administrative activities") could apply, but this is unclear because the AI system is used for individual cases.

ID 61

Enterprise function: Legal

Name: Evaluation of documents

Context:

The Civil Rights Corps (CRC), a nonprofit organization dedicated to combating systemic injustice within the U.S. legal system, faced a fact-intensive case with more than 300,000 documents to review. Faced with multiple defendants, a complicated set of facts, and many elements that needed to be corroborated, the fact-gathering process involved reviewing thousands of files to uncover evidence that would prove how the private parole system violated the constitutional rights of their clients.

AI system:

The investigative capabilities of an external e-discovery platform enabled the CRC team to discover mountains of evidence by quickly searching these files. They used a special story-building feature to

track the most critical documents from three defendants, collaborate virtually, and successfully prepare for depositions. By streamlining the e-discovery process from data upload and processing to search, review, and production, they were able to find meaningful information, bring hidden insights to light, and act on key evidence. As a result, they were able to attach 95 exhibits to their briefing motion for summary judgment. Since then, the team has used this AI solution in eight cases in seven states to search records, identify potential witnesses, and bring forth their unique stories.

What is unclear?

The AI system supports finding, linking and summarizing facts for ongoing processes and could thus fall under Annex III para 6 lit d) or para 8 lit a). In both cases, it is unclear whether the use is "on behalf of" a public body.

Annex II - Existing EU regulations

ID 101

Enterprise function: Research and development

Name: Product Development / Generative Design

Context:

The production of electric vehicles (EVs) presents many challenges. Although automotive companies are extremely optimistic about them - GM alone plans to launch at least 20 electric or fuel cell vehicles by 2023 - such vehicles are more expensive to produce. For GM, generative design could help solve these challenges by enabling lighter vehicles and a shorter supply chain. Electrification and autonomous vehicles will fundamentally change the automotive industry. Therefore, it will be critical for companies to be leaders in these highly technical areas in the future.

AI system:

In a recent collaboration and using generative design technology, GM engineers designed a new, functionally optimized seat bracket, a standard automotive part that secures seat belt attachments to seats and seats to floors. While the typical seat retainer is a boxy part consisting of eight pieces welded together, the software has developed more than 150 alternative designs that look more like a metallic object from outer space. GM's chosen design is made of a single stainless steel part instead of eight, is 40 percent lighter and 20 percent stronger than the previous seat mount.

What is unclear?

Electric vehicles are machines in the sense of the Machinery Directive (Directive 2006/42/EC), which is listed in Annex II Section A of the AI Regulation for risk classification. The seat belt mount could be considered a safety-critical component of the vehicle because the seat belt is attached to it. However, the AI system is only used to develop the bracket, but it is not part of the vehicle.

ID 89

Enterprise function: IT and Security

Name: Network Threat Analysis

Context:

A global Fortune 100 pharmaceutical company wants to optimize IT support for the Internet of Things (IoT), wearables and other non-standard devices deployed by its business units. To better serve its industry customers and reduce overhead costs, the company needs to securely and efficiently manage non-standard mobile devices across the enterprise.

AI system:

The IT team uses a unified endpoint management (UEM) tool from a third-party vendor to drive productivity and innovation across the organization and minimize overhead costs. End users can complete the entire self-service enrollment process, from registering to downloading the app, within 5 to 15 minutes. The tool helps keep the organization informed of potential endpoint threats and mitigations to prevent security breaches and disruptions. The solution provided clear visibility into and control over the company's global device inventory, which has grown to more than 80,000 company- and employee-owned devices and about 800 apps. It also helped the team save time and reduce costs by automating key configuration and support processes.

What is unclear?

This is a pharmaceutical company, so the affected end devices may include medical devices or in vitro diagnostic devices as defined in Annex II Section A of the AI Regulation, e.g. insulin pumps. It is unclear whether an AI system with the purpose of protecting internet-enabled medical devices or in vitro diagnostic devices from cyberattacks qualifies as a security component.

ID 82

Enterprise function: Logistics and supply chains

Name: Automatic inspection of economic goods

Context:

In industries such as logistics, damage and wear to operating equipment over time is commonplace. Using a camera bridge to photograph cargo trains, AI systems are able to successfully identify damage, classify the type of damage, and determine the appropriate corrective actions to repair them.

AI system:

First, cameras were installed along the railroad tracks to capture images of passing trains. The images were then automatically uploaded to an AI-based image store, where AI image classifiers identified damaged train components. The AI classifiers were trained on where to look for components in a given image and how to successfully identify such parts and then classify them into seven damage types. As more data was collected and processed, the AI system's visual recognition capability improved to an accuracy rate of over 90% in only a short period of time. The anomalies and damage detected by the system were sent to a workplace dashboard managed by maintenance teams.

What is unclear?

The rail system falls under Annex II Section B para. 5 and the purpose of the AI system is to identify and report damage to wagons. As undetected damage can be safety critical, it is unclear whether the AI system is a safety component in the sense of the AI Regulation.

Note: Because the AI system is classified as a high-risk system under Annex II, Section B (and not Section A), only Article 84 applies under Article 2 and not the high-risk requirements in Title 3.

ID 16

Enterprise function: Production and manufacturing

Name: Process control and optimization

Context:

Linde is a global supplier of industrial gases such as nitrogen, hydrogen, oxygen and many more, and is active along the entire value chain from production, processing and distribution to application. The operation and control of gas processing plants affect both productivity and costs, especially energy costs. Plant control includes the adjustment of individual compressors, pumps and turbines, heat exchangers and valves within a plant, but also the optimization of the overall system of a plant as a whole.

AI system:

Linde uses artificial intelligence to predict plant behavior and develop fine-tuned strategies to reduce energy consumption. The AI system is implemented using Deep Learning in combination with Reinforcement Learning. This means that the parameters of the plant and its components are mapped in a neural network, which then optimizes itself according to a predefined goal of the algorithm. For this purpose, machine learning engineers define this goal, the so-called reward function (e.g., a reduction in energy consumption), together with subject matter experts. The AI system is set up in a plant that produces oxygen and nitrogen and supplies them to a directly connected customer. Reliability and purity must therefore be stable at all times. Linde was able to refine the settings of individual components while the plant continued to run with stable performance.

What is unclear?

The KI system controls the energy supply of the gas processing plant, which presumably falls under the Machinery Directive according to Annex II Section A para. 1. It is possible that the plant also contains protective systems in potentially explosive atmospheres (Annex II para. 6 - Directive 2014/34/EU) or pressure equipment (Annex II para. 7 - Directive 2014/68/EU). It is unclear whether the energy supply of such a system is to be considered a safety-critical component.

Discussion: Causes of uncertainties

Based on the AI systems with unclear classifications, this chapter explores the underlying causes to formulate corresponding suggestions for improvement.

Critical infrastructure

Definition of "Critical Infrastructure": European or National Definition?

Discussion:

For the correct determination of high-risk systems in the area of critical infrastructure, it is essential which definition of critical infrastructure is used: the national or a European one? With the understanding that defining critical infrastructure is a national competence, the national definition of the German BSI was used in this analysis[1]. However, using the national definition hinders the AI Regulation's goal of strengthening the single market. In contrast, an AI system could be classified as high-risk in some Member States, contributing to fragmentation of supply and deployment.

The European Council's November 25, 2022 draft includes a definition of critical infrastructure with a reference to the resilience of critical entities Directive 2022/2557, after which the AI Regulation would follow the European definition (if the proposal is successful).

Proposal:

- Use of a European definition of critical infrastructure to strengthen the single market.
- Statements at national level, e.g. from the BSI in Germany, on which definition is to be applied until the (new) Directive 2022/2557[2] is adopted into national law.
- Create an overview for AI providers of how member states' definitions of critical infrastructure differ.

Asset types and thresholds: adjustments for AI?

Discussion:

The national definition of the German BSI specifies asset types and thresholds to identify so-called KRITIS companies. Looking at the AI systems under review, it is not clear whether the thresholds are applicable accordingly. For example, for an "Intelligent Transportation System"[3] , the "number of connected users or average users served in the service area" is 500,000 (see Intelligent Transportation Systems Act). Accordingly, would an AI system for route planning that also has safety-critical functions be considered a high-risk system?

1 https://www.bsi.bund.de/DE/Themen/KRITIS-und-regulierte-Unternehmen/kritis-und-regulierte-unternehmen_node.html
2 https://eur-lex.europa.eu/eli/dir/2022/2557/oj3
3 https://www.gesetze-im-internet.de/bsi-kritisv/anhang_7.html

Proposal:

- Review and, if necessary, adjustment of the types of assets and thresholds for AI systems by national bodies (e.g. BSI).
- Maintain and strengthen AI competencies of national bodies to support AI providers in applying the AI Regulation, e.g., through advice, guidelines, or in the context of AI Regulatory Sandboxes.

Definition of "safety component"

Discussion:

In some of the analyzed AI systems, it is unclear whether or not the AI system is a safety component. This finding is important because this is a necessary condition for classification as a high-risk system.

The draft of the European Council of November 25, 2022 contains in recital 34 examples of safety components in the area of critical infrastructure ("Examples of safety components of such critical infrastructure may include systems for monitoring water pressure or fire alarm controlling systems in cloud computing centers ..."), which are quite helpful.

To achieve climate goals (see European Green Deal), there is great potential for the use of AI, but where the classification is often unclear. For example:

- Prediction of traffic volume in cities and AI-assisted traffic flow, such as traffic light systems.
- Prediction of energy demand in buildings and AI-assisted heat generation in decentralized plants.
- Forecasting food demand and production to reduce waste.

Proposal:

- Cite other examples of AI safety components in all sectors of critical infrastructure e.g. water, health, food.
- Clarify whether a safety component can also be software only.

Employment

Definition of "Task"

Discussion:

In some cases, an AI system generates prompts for natural persons, but it is unclear whether this is a task in the meaning of the AI Regulation. Examples:

- Recommendations, such as for cooking instructions, for example, how to peel a pineapple.
- Instructions for navigation or route planning in road traffic.
- Orders that an employed person has to carry out, e.g. delivery services.

The draft European Council of November 25, 2022, includes in recital 36 that AI systems should be high risk for the purpose of task allocation because they have an impact on career development.

Accordingly, "tasks" seem to mean cases in which an employed person receives an instruction and which, if not fulfilled, has a direct, negative and personal impact on that person's professional development. However, cases where the recipient has the discretion to follow (or not), or for which there are alternatives, or when the suggestions are meant to inspire, would not be a "task" within the meaning of the AI Regulation. In particular, an AI-generated request is not a task, if failure to comply has no negative consequences for the person concerned.

Proposal:

- Addition of a definition for the term "task" to Article 3.

Interpretation of "work-related contractual relationships"

Discussion:

Another point with a need for interpretation in the context of task allocation by AI is the relationship between the operator (user/deployer) of the AI System and the person receiving tasks from the AI (affected person), because it depends on whether negative consequences are to be expected in case of non-compliance.

Variants of such a relationship include, for instance:

1. An employee of a food manufacturer uses a cooking chatbot for private purposes.
2. A call center employee receives customer requests assigned via AI.
3. A truck driver gets route suggestions from an AI.

Case 1 probably does not fall into the high-risk class because there are no negative consequences to fear in the event of non-compliance. In cases 2 and 3, classification as a high-risk system is possible because the company wants to achieve an increase in performance with the use of AI. The decisive factor here is what expectations and instructions the employer has expressed to the employee with regard to the AI system. If the employee has an alternative to the AI system and there is no threat of negative consequences, such AI systems may be low-risk.

Proposal:

- Clarify in which relationship constellations an AI system is to be classified as a high-risk system in the context of employment relationships.
- A possible concretization of the high-risk classification is, for example, when
  - the employer has authorized the use of the AI system.
  - the employer has instructed the employee to follow the recommendations or prompts of the AI system.
  - the employee faces negative personal consequences if he/she does not follow the AI system.

Law enforcement

"on behalf of"

Discussion:

The criteria in Annex III para. 6 apply to law enforcement authorities and to entities acting on their behalf, but it is unclear under which circumstances the second aspect applies. When is something done "on behalf of a law enforcement authority" and when is it not?

Examples:

- A financial institution that implements measures to prevent money laundering due to legal requirements
- An organization that uses an AI system to help witnesses in court present "better" testimony and evidence
- An attorney representing a client in court and using AI systems in preparation

In these cases, a public body (e.g., tax office, court) requests information from affected companies or individuals (sometimes by law), but in individual cases there is not necessarily a direct mandate.

The November 25, 2022 draft European Council includes the following exception in Recital 38:

"AI systems specifically intended to be used for administrative proceedings by tax and customs authorities as well as by financial intelligence units carrying out administrative tasks analyzing information pursuant to Union anti-money laundering legislation should not be considered high-risk AI systems (...)"

Proposal:

- Clarifying explanations to the AI Regulation under which circumstances the use of AI is "on behalf of a law enforcement authority," e.g., in recital 38
- Exhaustive listing of existing laws in the financial industry that may be an exception (in addition to the Anti Money Laundering Act).
- Build AI capabilities in the relevant bodies at national level (e.g. BaFin in Germany) to help companies classify correctly.

"criminal offence"

Discussion:

Annex III para.6 lit a) describes AI systems for predicting criminal offenses, whether as victim or offender. It is unclear what exactly is considered or meant as a criminal offense, because the term is not defined in the drafts of the AI Regulation. According to the German Criminal Code (StGB) § 12, any illegal behavior is considered a criminal offense. Many of the AI systems examined (with a view to use in companies) support activities that are regulated by law, e.g.road traffic, tax law, cybersecurity or labor law. Here, it must be clearly delineated under which conditions the AI system plays the central role, when it has a supporting function, and when the human is responsible.

Furthermore, it is unclear whether the national definition of “criminal offense” applies in each case or whether there is a uniform European definition.

Proposal:

- Definition of criminal offense, “criminal offense”, in the AI Regulation, or reference to an existing EU-level definition.
- Clarification under which conditions an AI system is supported and not to be classified as a high-risk system (cf. adminstrative tasks).

Evidence and facts

Discussion:

Appendix 3 para. 6 lit d) and para. 8 lit a) describe AI systems for evaluating or interpreting evidence and facts in the context of a trial. Diverse AI systems in this study support tasks in this area, such as searching large amounts of text for relevant evidence or reviewing contracts to gather specific facts (e.g., deadlines).

With regard to risk classification, it is unclear under which conditions a document (or other information) is to be classified as evidence or fact.

For example, is it required that a trial is already underway or are documents meant, that may become an evidence in the future, such as contracts that are reviewed by an AI application to meet deadlines.

Proposal:

- Definitions and delimitation for the terms “evidence” and “fact” in the AI Regulation.
- Explanatory notes in recitals 38 and 40

Annex II

"Safety component" in the AI regulation

Discussion:

Determining whether a product falls within the scope of the regulations in Annex II is usually straightforward. The question of whether an AI system is a safety component was unclear in some of the cases examined.

For example, when detecting damage on trains, the AI is neither part of the train nor of the railways. Image recognition in itself (cameras on bridges) is not part of a product or the "rail system" and as such does not pose a safety risk. But if the AI system does not function or functions incorrectly, consequential damage can occur.

In another example (not part of the study), an AI-powered smartphone app for predictive maintenance in elevators aims to predict, through the sensors in the smartphone, if there is damage to the elevator or if/when maintenance is needed. Again, the app is not part of the product (the elevator), but a failure or malfunction of the app can lead to damage.

Therefore, when identifying safety components, the determination of the system boundary is of central importance. In a narrow interpretation (focusing on the Annex II product), a supporting AI system, e.g. for predictive maintenance, may not be a safety component.

Note: The term "safety components" is not defined in the initial draft of the European Commission, but both the Parliament and the European Council have published corresponding proposals with the definition.

Proposal:

- Clear definition of "safety component" in the AI Regulation.
- Clarification of applicable system boundaries in the identification of safety components (e.g., in the recitals).

"Safety component" in the sectoral standards

Discussion:

The notion of "safety component" is central to risk classification and, in addition to the AI Regulation, the harmonized standards also play a major role because they specify the meaning and put it into sectoral context.

Ambiguities can arise here because there are already established standards that define the term "safety component", e.g. for medical devices in the context of essential performance (cf. IEC 60601-1:2022) or similarly in the automotive industry (in accordance with ISO 26262 and IEC 61508).

In addition, there are national laws and European regulations that contain similar definitions (e.g. BSI Law §2 (13) for Critical Components in Critical Infrastructure, Annex III in DIRECTIVE 2014/33/EU on Lifts and Safety Components for Lifts).

Proposal:

- When harmonized standards are determined by CEN/CENELEC, the different sectoral definitions are to be examined and evaluated and, if necessary, aligned.
- Competent authorities in Annex II sectors should be tasked to publish guidance documents for the identification of AI systems that are safety components in their respective sectors.
- Certification bodies are sensitive to sectoral variations in the definition of "safety component" and recognize them during conformity assessment.
- The variations in definition (via sectoral standards and national laws) are recognized in the AI Regulation, e.g., in recital 34

Redundant safeguards

Discussion:

According to the proposed definition of "safety component" in the EU Commission draft (April 2021), AI systems can be a safety component, if their failure creates a risk:

> *'safety component of a product or system' means a component of a product or of a system which fulfils a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property;*

In many safety-critical systems, redundant safety mechanisms already exist due to existing requirements (e.g., the cooling circuit in nuclear power plants, the power supply in hospitals).

Therefore, it is unclear whether an AI system with a safety function is not a safety component if there are redundant safety mechanisms in place. Example: In the AI-supported energy supply of a gas processing plant, there may be redundant systems that step in if the AI system fails or is faulty. In this case, would the AI system no longer be a safety component?

Proposal:

- Adjust the definition of safety components in the AI regulation with emphasis on cases where the AI system is the primary or sole safety component.

Recommendations

For politics

Area	Recommendations
Promote innovation	• Provide comprehensive guidance for the correct risk classification of AI systems, including clear instructions and examples, especially for AI in generic and industry-agnostic enterprise functions. • Binding and fast response to questions regarding unclear classification via a central European portal (to avoid different interpretations) e.g. in sandboxes. • Build up competencies, e.g. within the relevant authorities at the federal and state levels, to assist companies with the correct risk classification. • Implement information campaigns targeting all audiences (providers, operators, and individuals) to educate them on the requirements and obligations of the AI Regulation.
Reduce costs	• Standardization of definitions along national laws, European regulations and sectoral standards. Inconsistent definitions create redundant efforts without adding value. • Accelerate the development of standards and guidance documents that specify the requirements of the AI Regulation. Legal uncertainty delays the use of AI due to a fear of making mistakes. Accelerating the development or determination of harmonized standard is important because otherwise, after the end of the transition period, Article 43 para. 1 will require mandatory 3rd party certifications for all high-risk systems, which may lead to waiting times at certification bodies, i.e. market access.

Clarify requirements	Clarify the classification rules to ensure planning security for providers and users and to reduce the risk of incorrect classification. Definitions and interpretations (Article 3 and recitals): • Which definition (incl. asset types and thresholds) applies to "Critical Infrastructure", the European or the national one(s)? The European definition is more suitable for strengthening the internal market and in view of the cross-border nature of critical infrastructure. • Clarification of the definition of "safety component" (only hardware or also software?) and the system boundary to be considered (e.g. regarding predictive maintenance). • Clarification of the definition of "task" and the contractual relationship constellation in which negative consequences for the employee are expected. • Clarify when an AI system is used "on behalf of a law enforcement authority" and which exceptions are applicable (e.g. Anti-Money Laundering Act). • Clarification of the definition of "criminal offense" or reference to a definition at EU level. The same applies to the terms "evidence" and "fact".

For companies

Area	Recommendations
General aspects	• Focusing on AI systems with a high strategic value and a low risk class may make sense to keep the additional effort caused by the AI regulation low. • Business areas that tend to have many high-risk systems are human resources, customer service, accounting and finance, IT, and security. • Business areas that tend to have few high-risk systems are marketing and sales, purchasing, research and development, production and manufacturing, logistics, and supply chain. • Familiarize yourself with the AI Regulation and the classification rules (incl. the articles, recitals and annexes). Use available tools from appliedAI or similar points of contact.
In-house development of AI systems (Make AI)	• Perform initial risk classification at an early stage to avoid surprises later, because classification as a high-risk system increases cost and complexity and may affect the strategic value. • Involve diverse stakeholders in the risk classification, ideally bringing technical, legal, commercial, and user perspectives. • Have the result of the risk classification legally confirmed, especially in the case of larger investments, in order to increase planning certainty. In doing so, also consider possible variations of the AI system in the future. • Consider the risk class and applicable requirements across all phases of the AI lifecycle, especially for (foreseeable) changes.

Use of available AI systems (Buy AI)	• Ask the AI system vendor about the applicable risk class and whether it remains the same in your use case scenario. • Take into account who approves the use of the AI system, who works with it on a day-to-day basis, and who is affected by the results of the AI system. It is important that these people are informed at an early stage and are aware of their role. • Familiarize yourself with the requirements for users (“Users” or “Deployers”) of the AI Regulation. • Discuss changes to the deployment scenario with the AI provider in advance, as changes may cause reclassification.

Limitations

We have conducted this study to the best of our knowledge and conscience and therefore explicitly point out the following limitation:

Focus on AI in the enterprise

The AI systems analyzed are all taken from the corporate context, i.e. AI in other application areas, such as for specific industries (e.g. medicine, aerospace, automotive) or sectors (e.g. education, public administration, healthcare), are not included. Thus, the results are not representative for the totality of all AI applications, but they give a very good and broad overview of AI in functional areas of enterprises.

Significance of the examined AI systems for companies in Europe

The AI systems considered are currently in use, but not only in Europe. Therefore, we cannot make any statement as to whether and to what extent the selection of AI systems is representative of AI in European companies.

Limited information about AI systems

Descriptions of AI systems were limited and in some cases the lack of detail was a reason for an unclear classification. With more information, the proportion of unclear cases would be possibly lower. This observation shows that comprehensive details about the AI system need to be known for an unambiguous classification.

Amendments to the AI Regulation

This study was prepared during the ongoing negotiations of the AI regulation. We took special care to apply the same standard to all AI systems and to reflect divergent rules from recent drafts and to indicate them as such. Future changes may result in different classifications.

Possible errors in the analysis

AI regulation is a comprehensive and complex set of rules, and AI is a complex and multi-faceted technology. Both are continuously evolving. The authors have dealt extensively with both topics and there have been various review cycles to check the study with lawyers as well as experts from the EU institutions. Nevertheless, it cannot be ruled out that errors may have crept in while drafting process this study.

Authors

Dr. Andreas Liebl

Dr. Andreas Liebl, Managing Director of appliedAI, is responsible for the appliedAI initiative. He serves as an expert regarding innovation and commercialization and as steering committee member in the Global Partnership on AI next to other advisory roles as in the Plattform Lernende Systeme or the KI Rat Bayern. In previous roles, he was a managing director of UnternehmerTUM, one of the largest innovation and entrepreneurship centers in Europe, he worked for McKinsey for five years and did his PhD at the Technical University of Munich.

a.liebl@appliedai.de

Dr. Till Klein

Dr. Till Klein, Head of Trustworthy AI at appliedAI, is driving the acceleration of AI by co-creating methods, tools and insights in the field of AI Regulation and Governance to enable compliance and trust by adopters. He is a member of the OECD.AI and he has several years of industry experience in the regulatory roles, including Medical Devices, Drones, and as Quality Management System Auditor. Till is an Industrial Engineer and did his PhD on the evolution of collaboration networks in the context of technology transfer.

t.klein@appliedai.de

Publisher

initiative for applied artificial intelligence

appliedAI Initiative GmbH
www.appliedai.de

Freddie Mercury Street 5
80797 Munich, Germany

The appliedAI initiative is Europe's largest initiative for the application of cutting-edge, trusted AI technology with the goal of making European industry shapers in the age of AI, creating a world we want to live in. appliedAI acts as both an enabler and an innovator.

Bavarian State Ministry for Digital Affairs

Bavarian State Ministry for Digital Affairs
www.stmd.bayern.de

Oskar-von-Miller-Ring 35
80333 Munich, Germany

Acknowledgement

The authors would like to thank Dr. Tim Christiansen and Bianca Rabl from the Bavarian State Ministry for Digital Affairs for their excellent cooperation in the implementation and publication of this study.

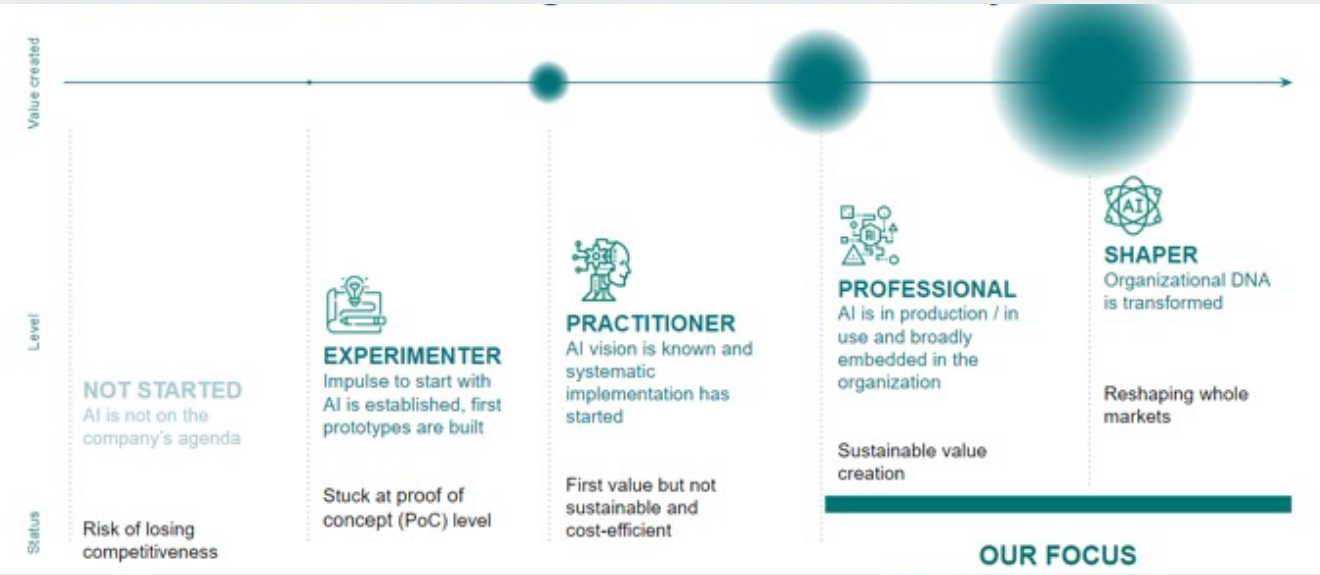
Sincere thanks also go to attorney Dr. David Bomhard and attorney Jeannette Gorzala for their feedback and suggestions.

We would also like to express our gratitude for the very constructive exchange with employees of the European institutions (Commission and Parliament).

Many thanks to the team at appliedAI who helped in the preparation of the study, especially Susanne Klausing and Manuel Jimenez Medira.

Information about appliedAI

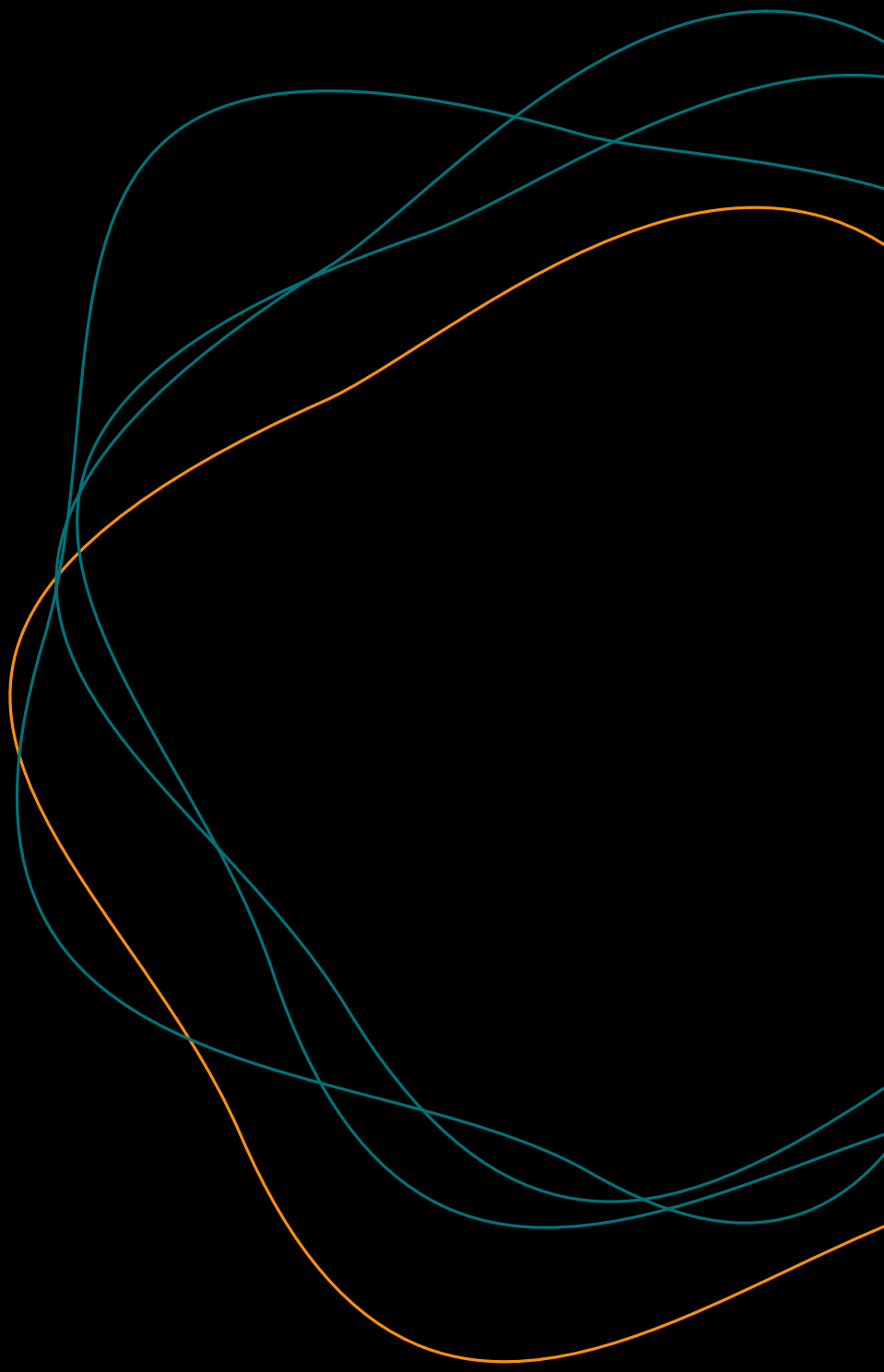
The appliedAI initiative was launched in 2017 and currently employs over 80 people with different backgrounds in the field of Artificial Intelligence (AI). Its goal is to turn European industry into shapers in the age of AI, creating a world we want to live in. In the global race for technology leadership, this goal can only be achieved by working together and learning from each other. In doing so, we focus on accompanying companies into professional AI application, as this is the only place where added value is truly created.



appliedAI works with companies that share a mentality of collaboration and openness through partnerships to create, access, and share unique knowledge for the application of trusted AI. Furthermore, the initiative supports AI transformation with solutions and services as well as comprehensive programs to accelerate AI adoption.

In this context, appliedAI is committed to the competitiveness of the European industry while complying with future regulatory frameworks and is carrying out concrete activities such as the development of an AI risk classification tool and the establishment of an MLOps infrastructure, including tools and processes, for compliance with the AI Act.

Companies interested in cooperation are welcome to contact us.



White paper

appliedAI Initiative GmbH
www.appliedai.de

Freddie Mercury Street 5
80797 Munich, Germany

initiative for
applied artificial
intelligence