Universiteit Leiden
The Netherlands

Opleiding Informatica

CompAI: A user-friendly software tool to support the CapAI procedure for AI Act conformity assessment.

Camiel van Schaik

Supervisors:
Prof. dr. ir. J.M.W. Visser
Dr. P.W.H. van der Putten

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

26/06/2023

Abstract

Background: Artificial intelligence has recently experienced a surge in not only its capabilities but in its application as well. This has resulted in a growing need for systems that ensure the safety, robustness and fairness of AI implementations. The European Union's Artificial Intelligence Act (AI Act) is the most extensive piece of upcoming legislation which offers a set of harmonised rules to aid in the design of trustworthy AI systems. There are different frameworks and industry standards that try to assure the deployment of responsible AI. However, many of these systems do not ensure compliance with the AI Act. CapAI is a governance tool that focuses on conformity with the AI Act. It takes the entire AI lifecycle, from design to retirement, into account and defines and reviews current practices to assess each stage of the lifecycle. CapAI offers a formal way to conform to the AI Act. However, the industry has yet to come up with a way to implement the CapAI procedure in a user-friendly and clear manner.

Objective: This thesis investigates a way to streamline the usage of CapAI to make it easier to implement the framework into the AI lifecycle. The proposed solution is a software solution called CompAI which guides users through the CapAI procedure and gives insight into the overall compliance of the AI system and organisation.

Method: CompAI documents any information necessary to comply with the AI Act. This entails the documentation of the execution of the internal review protocol (IRP) and the visualisation of the summary datasheet (SDS) and external scoreboard (ESC). Furthermore, CompAI gives a clear insight to key actors about the conformity of their AI systems through visualisations. It also guides these actors through all of the CapAI principles during the entire lifecycle of the system. Interviews with industry professionals are conducted to measure the usability and effectiveness of the tool.

Results: The CapAI procedures are fully implemented in the proposed software solution called CompAI. The open-source system leads the user through the CapAI procedure and outputs an IRP, SDS and ESC. Industry professionals have reviewed the proposed solution. The review has shown that CompAI has high perceived usefulness.

Conclusion: CompAI shows in a user-friendly way how users can execute the CapAI procedure. The review has shown that the system can be improved by expanding the system to speed up the IRP and take away the need for detailed knowledge about the AI Act.

Contents

Chapter 1

Introduction

1.1 Motivation

On 21 April 2021, the European Commission published a proposal to regulate artificial intelligence in the European Union, the AI Act [1]. The regulation should harmonise existing regulations and ensure that AI systems are safe and respect existing laws and fundamental rights [2]. The AI Act comes with a plethora of requirements, one of which is the conformity assessment. CapAI is a procedure created by researchers from the University of Oxford for conducting these conformity assessments. However, guidance on how to utilize this procedure is necessary for companies to harness its full potential [3].

1.2 Research goal

This thesis explores the possibilities for implementing the CapAI procedure into a software solution. This is done to simplify the execution of conformity assessment of AI systems in line with the EU Artificial Intelligence Act and to aid the communication around the AI Act within project teams. The proposed solution should possess a high perceived usefulness according to industry professionals.

1.3 Research approach

The methodology chosen for this thesis is Design Science Research [4]. Design science research (DSR) is an approach that aims to develop and evaluate innovative solutions to real-world problems by creating and testing artifacts. These artifacts can be tangible (e.g., software applications, algorithms) or intangible (e.g., design principles, theories). For the purpose of this research, we created 2 tangible artifacts: The maturity model and the CompAI software tool. The Design Science Research Process consists of six activities in a nominal sequence, Figure 1.1 presents this process graphically. This thesis handles the DSRP from problem identification to evaluation. As of writing this thesis, the AI Act is still going through the legislative process. Therefore, we decided to take as subject of this research the 2021 proposal of the AI, since this provides us with the most stable data to base our research on. The only deviation from this is Section 2.2 which references the European Council's general approach on the AI Act from 6 December 2022 [5].

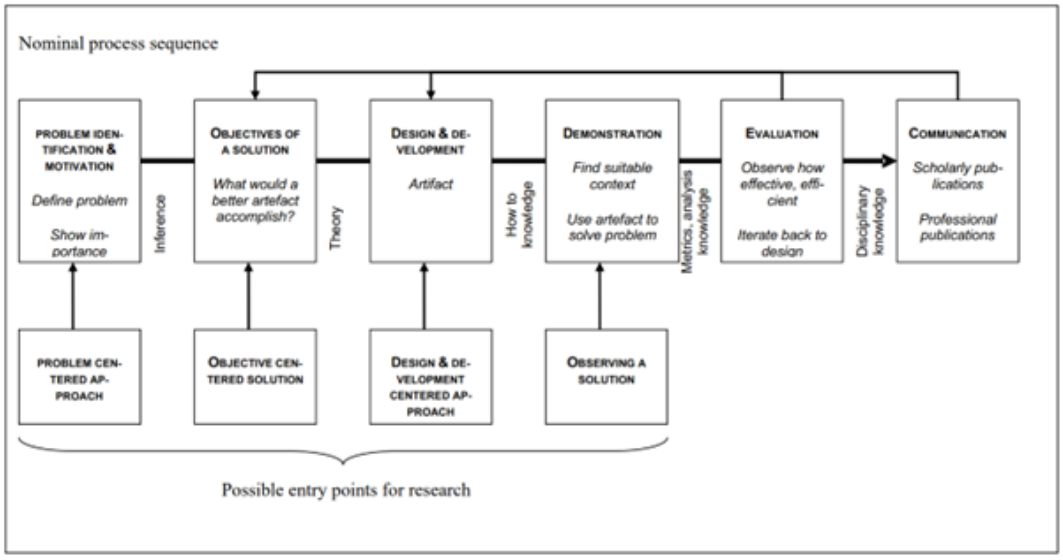


Figure 1.1: The Design Science Research Process (DSRP) model. Outlining the different stages of DSR and their entry points. Source: [4, p. 11]

1.4 Structure

First, in Chapter 2 we will dive into the background of this research. We will take a look at the AI ethics landscape and see how this landscape results in the need for AI regulation. Then, we will discuss the intricacies of the AI Act to see what this new regulation will mean for AI providers. Moreover, we will dissect the CapAI framework to see what features will need to be implemented into CompAI. Chapter 3 describes the design principles used to develop both the CapAI maturity model and the CompAI software tool. Subsequently, Chapter 4 lays down the review process and the herein-acquired results. Chapter 5 uses these results to assess the the usefulness of CompAI and define areas for future research. The thesis is concluded in Chapter 6.

Chapter 2

Background

In this chapter, we will discuss the current state of AI ethics and how this influences the need for regulation. Furthermore, we will take a deep dive into the AI Act and how it corresponds with this aforementioned need. To do this we will explore the Act through the lens of AI organisations to explain how the legislation will affect them and which new requirements await AI providers. Subsequently, we will discuss the criticism that the AI Act proposal has faced and weigh in on methods to improve the new regulation. Then we will conclude this Chapter with a review of the CapAI framework. Here we will look at the different tools provided by CapAI to ease the adoption of AI Act principles for organizations.

2.1 The landscape of AI and Ethics

Artificial Intelligence is increasingly gaining more relevance in today's society. The amount of publications in the field of AI has doubled from 200.000 in 2010 to almost 500.000 in 2021 [6]. However, most progress is not found in academia but rather in industry, where most significant machine learning models are currently produced [6]. The global artificial intelligence market was valued at USD 428 billion dollars in 2022 and has been projected to grow to over USD 2.000 billion by 2030 [7]. AI systems are becoming available to more consumers and are being intertwined with popular products like Office 365, Bing, Snapchat and the Chinese e-commerce platform Alibaba.

Although these new developments support the belief that AI has "the potential to bring significant benefits to [society]" [8]. It has become a common acknowledgement that "AI technologies yield powerful advances but also can threaten [societal] values and fundamental freedoms if they are not developed and deployed responsibly or if they are misused" [8]. With this acknowledgement comes the call for the regulation of AI to prevent these detrimental effects.

Regulation of AI can be achieved in two ways, either by self-regulation from within the industry itself or by legislation enacted by governments. Both of these methods have their trade-offs. "Self-regulation is more desirable than government regulation if the degree of asymmetric information between the public regulator and private industry is larger than the size of the monopoly distortion and externalities from the industry to society. An optimal mechanism consists of both self-regulation and government regulation" [9]. Self-regulation is executed with a plethora of methods and more AI ethics tools are being developed. [10] defines three categories of AI ethics tools to create more structure within this landscape.

2.1.1 Impact Assessments

In the first place, there are impact assessments. This can be "a type of fact-finding and evaluation that precedes or accompanies research, or the production of artifacts and systems, according to specified criteria" [10, p. 407]. However, in practice impact assessments are used to assess systems after they have been deployed as well. "These assessments are shaped by notions of relevance (what is important to society and which phenomena are worthy of attention), evidence (identification of causes and effects), and normative claims (what is good, acceptable or tolerable)" [10, p. 407].

2.1.2 Technical and Design Tools

Second, the paper distinguishes technical and design tools. These tools typically originate from within the AI/ML community itself. These can be computational. Providing metrics to benchmark ethics principles such as fairness and bias. Moreover, they can consist of awareness workshops to raise awareness of AI ethics and implement them further into the design process [10].

2.1.3 Auditing Tools

At last, auditing tools are defined. This is the process of verifying the artifacts that record decisions, systems and processes against standards, legislation or other metrics. Audits need to be conducted independently by a third party. The goal of an audit is to create transparency for "a broader range of stakeholders beyond the entity or process in question" [10, p. 408].

2.1.4 Shifting the Focus of Self-regulation

The question remains if self-regulation provides adequate measures to ensure ethical AI. [11] analyzed a corpus of ethical AI principles and guidelines until 2019. The research states that both the private and public sectors had published a nearly equivalent proportion of documents. This would indicate that both parties are concerned with the ethical challenges of AI. However, while further investigation indicates that there is convergence on the importance of transparency, responsibility, non-maleficence, and privacy within the AI lifecycle, there was significant divergence in four major factors. These factors were: how ethical principles are interpreted; why they are deemed important; what issue, domain, or actors they pertain to; and how they should be implemented. This divergence indicated that stakeholders have different interests which are reflected in their guidelines on ethical AI. This calls for a harmonisation of AI ethics and a shift from the mere formulation of principles to actual ethical AI practice.

The views of these stakeholders on AI ethics are explored further in [12]. The report asked 602 experts in the field of AI to give their opinion on the question: "By 2030, will most of the AI systems being used by organisations of all sorts employ ethical principles focused primarily on the public good?" 68% of the respondents answered: "NO, ethical principles focused primarily on the public good WILL NOT be employed in most AI systems by 2030." The most predominant factor mentioned throughout the paper is the skewed prioritisation by AI developers. Respondents noted that effectiveness has been driving AI innovation, not ethics. Furthermore, the paper states that global competition, especially between China and the U.S., is causing an arms race that pushes the prioritisation of effective AI even further. The fact that the aforementioned countries define ethics in

different ways does not change this situation for the better. The paper describes a lack of incentive for corporations to correct this prioritisation of efficiency as described above. As discussed at the beginning of Section 2 most developments arise in the private sector. Therefore it is paramount that businesses experience benefits from creating an ethical AI lifecycle.

One way to provide this incentive is through certifications. Currently, there is no standardized and widely accepted certification for AI ethics. The ecosystem of AI ethics certifications mostly consists of stand-alone programs developed by individual government bodies and institutions [13]. However, a more standardized certification program could reduce information asymmetries by causing transparency in the ethics principles implemented into the system and the development process [13]. Furthermore, corporations will be incentivized to achieve certain ethics standards if these certifications are valued by their customers [13]. This way the AI ethics landscape could achieve the harmonisation it needs and shift from principles toward the actual practice of ethical AI.

2.2 Artificial Intelligence Act

The European Union is developing the Artificial Intelligence Act (AI Act) [1] as a reaction to the need for harmonisation of AI ethics. With this new legislation, the EU tries to address this need for harmonisation in a way similar to that of the General Data Protection Regulation (GDPR) [14] privacy law enacted in 2018. The AI Act specifies four objectives to do this [1]:

- ensure that AI systems are safe and respect existing laws and fundamental rights [2];
- ensure legal certainty;
- enhance governance and effective enforcement of existing law on fundamental rights and safety requirements;
- facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

2.2.1 The definition of AI

The AI Act takes a hybrid approach to defining what artificial intelligence entails. The regulation specifies both a broad definition as well as special categories and use cases for AI. The broad definition of an artificial intelligence system as defined in Article 3 of the AI Act: "software that is developed with one or more of the techniques that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" [1, p. 39]. The techniques mentioned are:

- "Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning";
- "Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems";

- "Statistical approaches, Bayesian estimation, search and optimisation methods" (Not part of the definition anymore since December 2022 [5]).

The AI Act has adopted a clause that gives the EU Commission the power to update this list of techniques [1, art. 4]. This is done to make the regulation future-proof and up-to-date with market and technological developments. An example of this has already been shown after The EU Council adopted its common position ('general approach') on the AI Act [5]. This document excluded statistical approaches from the definition of AI to be able to make a clear distinction between AI and simpler systems. What the impact of this will be in practice remains the question.

2.2.2 Territorial scope of the AI Act

The scope of the AI Act again exhibits some resemblance with the GDPR [14]. The territorial scope of the regulation can be summarised as [1, art. 2]:

- Providers who place on the market or into service AI systems in the EU;
- Users of AI systems located within the Union;
- Providers and users of AI systems where the output of the system is used in the Union.

Notable about this is the expansive territorial jurisdiction of the AI Act. Not only providers and users within the Union will be affected but those outside it as well. When these AI systems or their output is used within the EU the AI Act will apply, just like with the territorial scope of the GDPR [14, art. 3]. This points out the European Commission's inclination to de facto externalise its laws to apply outside its borders. The scoping of the AI Act will make it likely for the regulation to become a standard for AI ethics [15]. This is also called the 'Brussels Effect' [16], the global adoption of EU regulations through market mechanisms.

2.2.3 The four layers of AI

The AI Act orders AI systems using a risk-based approach [1, p. 7] and handles them with a layered enforcement mechanism [17]. This means that systems with minimal risk are met with fewer obligations than those with a high risk and applications with an unacceptable risk are even banned. Figure 2.1 illustrates the structure of the four layers, the associated AI Act articles, what key obligations they hold and examples of systems within these layers. In descending levels of risk, we will go through the four layers identified by the AI Act and discuss the criteria of each category.

Unacceptable risk

Systems that fall under the category of Unacceptable Risk will be prohibited with the enactment of the AI Act. "The criterion for qualification as an Unacceptable Risk AI system is the harm requirement" [17, p. 3]. Therefore, the AI Act describes these types of systems as: "AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights" [1, p. 12]. More specifically, the AI Act defines four categories of such systems, these can be summarised as [1, art. 5] (amendments from the EU council 'general approach' are added in brackets):

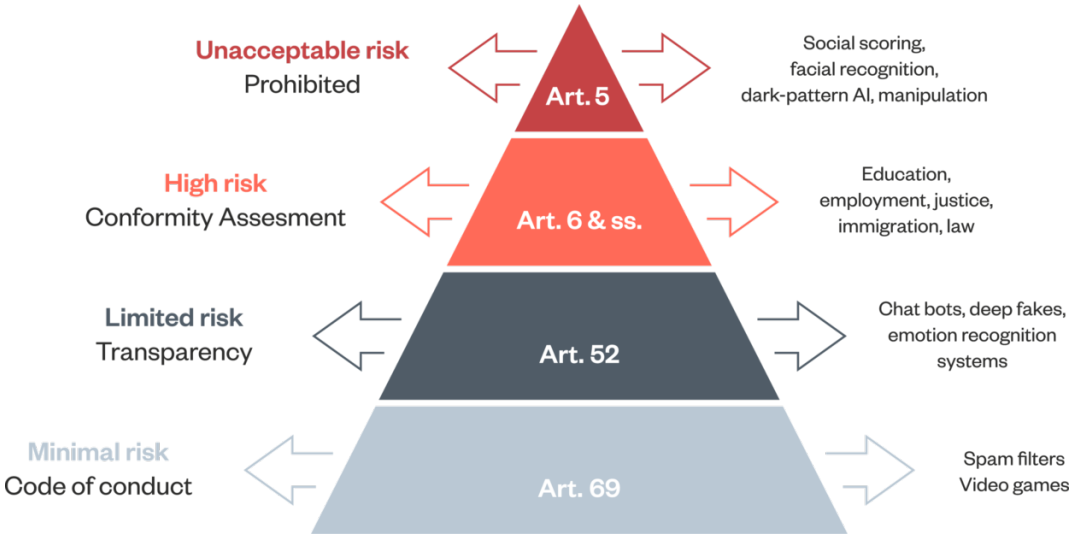


Figure 2.1: The four layers of AI classification as defined by the risk-based approach of the AI Act. Left, the name of each layer along with the main requirement for that layer are shown. Center, the AI Act article that outlines the provision of this layer. Right, examples of use cases that are considered part of each layer. Source: [18]

- Subliminal techniques that distort a person's behaviour that may cause physical or mental harm;
- Systems that exploit vulnerabilities of specific groups of persons due to age, disability (or "social or economic situation" [5]) to distort a person's behaviour that may cause physical or mental harm;
- Social scoring systems in the public sector (and by "private actors" [5]). Where the scoring leads to detrimental or unfavourable treatment of natural persons either, in social contexts unrelated to the contexts in which the data was collected, or that is unjustified or disproportionate to their social behaviour;
- real-time remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement.

Moreover, the last category can only be used after law enforcement authorities are exceptionally allowed to use such systems [5] and if it is strictly necessary for one of the following goals [1]:

- "the [ ] search for specific potential victims of crime";
- "the prevention of a specific, substantial and imminent threat[ ] physical safety of natural persons or of a terrorist attack";
- "the detection, localisation, identification, or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State".

High-risk

High-Risk AI systems pose a severe threat to the fundamental rights of individuals and are therefore subject to the strictest regulations under the AI Act. The systems that are part of this layer can be described by the following categories [1, art. 6]:

- "AI systems intended to be used as a safety component of a product, or itself a product, which is already regulated under the New Legislative Framework (NLF) [19] (e.g. machinery, toys, medical devices) and other categories of harmonised EU law (e.g. boats, rail, motor vehicles, aircraft, etc.)" [18].
- AI systems listed in any of the following areas:
  - **Biometric** 'real-time' and 'post' remote **identification and categorisation** of natural persons;
  - **Management and operation of critical infrastructure** safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity;
  - **Education and vocational training**, to determine access to education or professional training;
  - **Employment, workers management and access to self-employment**, for recruitment or performance and behavior evaluation.
  - **Essential private and public services**, for automated welfare, benefit systems, credit scoring and first respond services;
  - **Law enforcement**, systems that may pose a risk to people's fundamental rights, such as deepfake detection, pre-crime detection and crime analytics;
  - **Migration, asylum and border control management** for example to verify the authenticity of travel documents;
  - **Administration of justice and democratic processes** to assist a judicial authority in researching interpreting and applying facts and the law.

As discussed in 2.2.1 with the definition of an AI system, the EU Commission again has the power to add AI systems to the high-risk category if used in the aforementioned areas.

Limited risk

The next category of AI systems is that of limited risk. [1, art. 52] specifies three different types of systems that fall under this category.

- Chatbots;
- Systems for emotion recognition and biometric categorisation;
- Systems generating deepfake or synthetic content.

Minimal risk

At last, there is the category of minimal risk. These systems don't process personal data or do not affect any individual directly or indirectly like spam detectors, or AI in video games. as of writing this thesis, these systems are not subject to any strict requirements. However, in memorandum 5.2.7 [1, p. 16] the AI Act does encourage Providers of these systems to regulate them through voluntary codes of conduct.

2.2.4 Requirements

There is a multitude of legislative requirements for the different risk categories of AI under the AI Act. In this section, we will go over each category and discuss the impact the AI Act has on each of them. It should be noted that the AI Act does bring other legislative measures, like establishing the European Artificial Intelligence Board[1, art. 56]. However, in this paper, we will primarily focus on the AI Act from an AI provider standpoint and not dive deeper into these aspects of the AI Act.

High-risk

As discussed in Section 2.2.3 High-risk AI systems are subject to the most invasive regulations under the AI Act. The Act defines the following essential requirements for these systems.

- **Risk management system (Article 9):** implementing processes to identify, analyze and mitigate risks during the entire AI lifecycle;
- **Data/data governance (Article 10):** Data quality should be ensured by implementing measures surrounding training data, data preparation and biases.
- **Technical documentation (Article 11):** Up-to-date documentation should be published before the system is placed on the market or put into service
- **Record-keeping (Article 12):** The system should be designed to automatically log events to ensure traceability of the systems' functioning;
- **Transparency (Article 13):** The system shall be accompanied by instructions for use which include characteristics, capabilities and limitations of the system;
- **Human oversight (Article 14):** It should be possible for natural persons to oversee the system by understanding its workings and output and being able to intervene;
- **Accuracy, robustness and cybersecurity (Article 15):** The system should demonstrate to be accurate and resilient to errors, inconsistencies and cyber-attacks by accuracy metrics and fail-safe plans;
- **Quality management system (Article 17):** The provider of the system shall have policies, procedures and instructions in place to ensure quality through the entire lifecycle;
- **Post-market monitoring (Article 61):** The provider should have a system in place to analyze the system's performance.

Furthermore, High-risk systems will need to bear the CE marking to indicate conformity with the regulation before being put on the EU market [1, art. 16]. The CE marking can be acquired by performing a conformity assessment [1, art. 19]. The procedure for this conformity assessment is dependent on the type of AI system. As discussed in Section 2.2.3 the High-risk category distinguishes between systems that are already regulated under the NLF or other categories of harmonised EU law and those that are not. In the case of these already regulated systems, there will be mandatory external assessments from a third-party "notified body" [1, art. 43.4]. The same will be the case for AI systems used for biometric identification or categorisation of natural persons [1, art. 43.1]. Unless some type of technical harmonised standard is made for these systems, which will make an external assessment redundant [18, p. 20]. For the other categories of High-risk AI systems, it will be sufficient to conduct a self-assessment focused on the same requirements without the involvement of a third party to achieve the CE marking [1, art. 43.1].

furthermore, the conformity assessment will assess risks around the aforementioned requirements. Providers will have to identify these risks and formulate mitigating measures. Residual risks will have to be communicated to users whenever these risks cannot be eliminated. When providers can justify that they comply with these requirements the system will be able to bear the CE marking and be freely distributed in the EU [1, annex VI, VII].

When the system has gone on the market it is paramount that, despite modification, learning or changing usage, it stays compliant with the essential requirements. The post-market monitoring system, established by providers in conformity with the essential requirements, should notify providers and deployers of these systems about any new risks, serious incidents or malfunctioning [1, art. 61]. If any incidents or malfunctions are detected they should be reported to the Market Surveillance Authority (MSA) within 15 days. These MSAs are the national supervisory authorities under the AI Act [1, art. 62]. Member states will have to establish these bodies or can in some cases delegate these roles to Data Protection Authorities [1, art. 59]. Whenever MSAs are unable to effectively execute their task or are in need of advice they will be able to turn to the EU AI Board which will be established under the AI Act [1, art. 56].

To accommodate both MSAs and the EU AI Board to keep track of all High-risk AI systems there will be an AI database which will be controlled by the EU AI Board [1, art. 60]. Every provider will need to register their High-risk system upon market entry. The database should provide a better understanding of the overall AI landscape and ease governance and control of these systems by the governing bodies.

Limited risk

Limited risk AI systems are subject to a minimal set of transparency requirements [1, art. 52]. Providers of chatbots must ensure that the system is designed such that users are not interacting with a human but rather a machine. In contrast, the AI Act denotes that **users** of systems for emotion recognition, biometric categorisation, deepfakes or synthetic content should disclose to persons exposed to them that these systems were used.

Minimal risk

The Act does not propose any requirements for these systems. However, it does encourage the drawing up of voluntary codes of conduct [1, art. 69]. The act specifically mentions these codes of

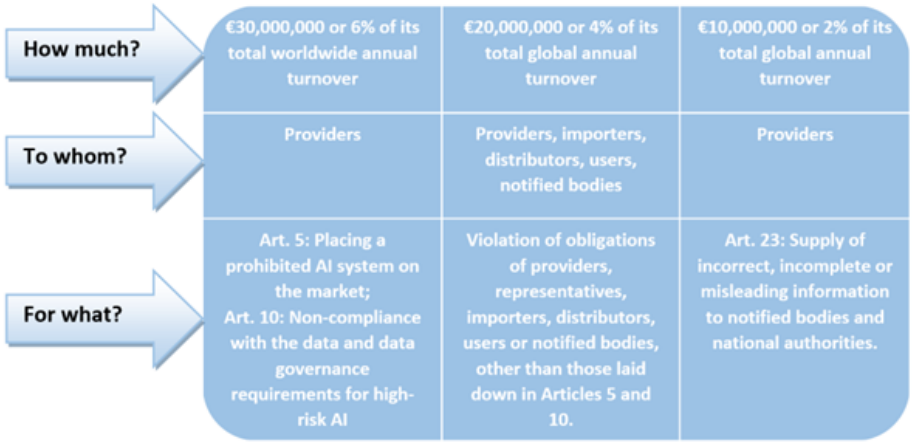


Figure 2.2: The penalties under the AI Act.

conduct could be focused on topics such as: "environmental sustainability, accessibility for persons with a disability, stakeholders' participation in the design and development of the AI systems and diversity of development teams on the basis of clear objectives and key performance indicators to measure the achievement of those objectives" [1, p. 80].

Penalties

Non-compliance with the requirements described in Section 2.2.4 will be met with serious sanctions. [1, Title X] bestows MSAs with the power to fine organisations when they violate the regulation. The AI Act groups these violations into three major themes. Figure 2.2 shows these different themes of violations, who is held responsible for them and the maximum administrative fines defined by the AI Act [1, art. 71].

Each Member State is able to define further rules within the confines of Title X [1, Art. 71(1)]. For instance, the Member States should lay down rules on administrative fines for public authorities and bodies established in that Member State [1, Art. 71(7)]. The AI Act also emphasises that the decision process for the amount of the administrative fine should be made on a case-by-case basis [1, Art. 71(6)]. Specifically, MSAs should take into account the following criteria when calculating fines [1, Art. 71(6)]:

- "The nature, gravity and duration of the infringement and of its consequences;"
- "Whether administrative fines have been already applied by other market surveillance authorities to the same operator for the same infringement;"
- "The size and market share of the operator committing the infringement."

High-risk AI systems seem to be most likely affected by these penalties since most obligations under the AI Act focus on these systems. The severity of these penalties should incentivize organisations to operate conforming to the legislation.

2.2.5 Criticism

The AI Act has faced a plethora of criticism and discussion during its legislative process. In this section, we will discuss some of the key points of discussion around the proposal and give our own insights into these issues.

General purpose AI

If we look at the risk-based categorization of AI as described in Section 2.2.3 it becomes evident that General Purpose AI does not necessarily belong in any of the high-risk system groups. General purpose AI systems have a multitude of possible uses depending on the context in which they are operated. Most times it is the user of the system that decides the purpose for which the AI is used [20]. Examples of these types of AI are large language models such as OpenAI's ChatGPT. Since these types of AI do not have a set purpose it is likely that providers of such systems will not be obligated to comply with the requirements for High-risk systems [21]. Furthermore, many of these models are integrated by different deployers than the original provider into downstream applications. Making only use of the output of these AI as a service capabilities [22] the deployers would be able to integrate general-purpose AI without modifying it. As of writing this thesis, the AI Acts language could result in these deployers not being deemed providers [21], as providers are defined as an entity who "develops an AI system or that has an AI system developed [ ] or [puts] it into service under its own name or trademark" [1, art. 3]. This definition could leave room for loopholes in some cases. Therefore, this secondary deployer would not be liable for certifying the system against the Act's requirements.

With this in mind, it would be beneficial for the robustness of the legislation to modify the categorization of the AI Act. The primary flaw we see in the current risk-based approach is that the risks are tied to certain use cases of AI. However, the development of the AI landscape can be unpredictable as seen with the uprise of large language models [20]. The question will be if the legislative process after the enactment of the AI Act will be fast enough to keep up with these turbulent changes and update the Act accordingly [21]. Hence, there might be a need to shift the categorization of high-risk systems back to its original purpose: to address all AI systems with great risk to the freedom and rights of natural persons. Using this formulation for High-risk AI systems alongside the use cases already adopted in the AI Act ambiguity can be prevented while at the same time creating legislation independent from technologies or narrow use cases. This would, however, make it necessary to create a standardized risk assessment for AI systems with which providers can assess their product to determine the category [20].

No subject rights

Another point of critique is the lack of consumer rights within the AI Act [23]. Compared to modern data protection law [14, art. 80] the AI Act does not provide subjects of AI systems the legal right to sue a provider or user for failure to comply with the Acts requirements. This could cause problems when regulators turn out to be ineffective in the enforcement of the act. Due to this lack of bottom-up force to hold regulators accountable, individuals whose fundamental rights are affected could be left powerless.

Severe impact on SMEs

A survey analysing the AI Act's impact on start-ups in Europe [24] has pointed out that the AI Act's initial impact assessment [25] might not be accurate. The survey "found many European startups [ ] concerned with the current direction [ ] of the AI Act, as 33-50% of respondents would see their technology potentially falling into the high-risk classification of the [ ] proposal". This would be a significant deviation from the envisaged 5-15% in the AI Act's initial impact assessment [25]. The survey points out that this could lead to a stagnation of AI innovation in the EU. The costs of compliance for SMEs are also expected to make a severe impact on the market [26]. "Compliance costs are likely to exceed those incurred by the GDPR threefold" [27]. Again it is likely that the cost of compliance will be higher than estimated in the EU impact assessment. The initial assessment predicted that A European SME that deploys a high-risk AI system will incur compliance costs of up to €160.000 [25]. However, a more recent study by the Center for Data Innovation estimated compliance costs of up to €400,000, which would cause profits to decline by 40 percent [26].

National security exception

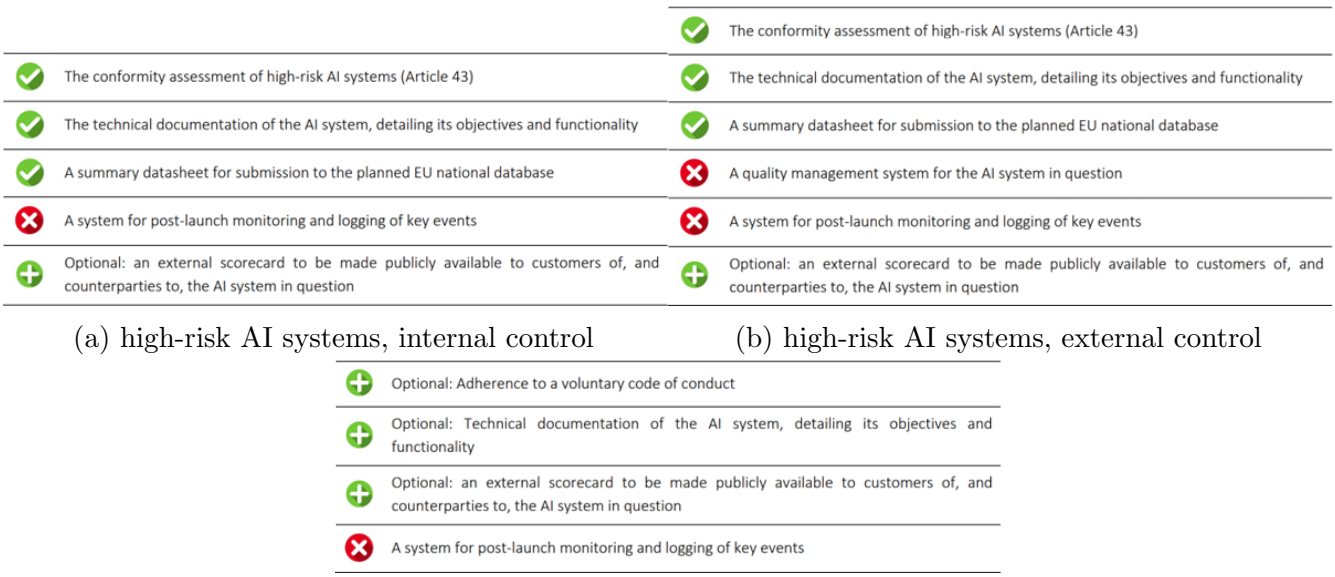
With the coming of the December 2022 EU Council general approach [5] exceptions have been made regarding the AI systems used for national security, defence, or military purposes. The general approach states that these systems are outside the scope of the AI Act. Human rights advocates are warning that these exclusions can pose severe risks to people's freedom and rights [23]. By allowing invasive AI systems, for example, social scoring or biometric mass surveillance systems, under the guise of "national security" the act could play into the hands of autocratic governments [23].

2.3 CapAI

In this section, we will discuss the compliance framework CapAI, designed by researchers at the University of Oxford [28]. CapAI is a "conformity assessment procedure for AI systems, to provide an independent, comparable, quantifiable, and accountable assessment of AI systems that conforms with the proposed AIA regulation" [1, p. 3]. CapAI's primary function is to act as a governance tool to guarantee and prove the development and management of trustworthy AI. This is done by providing "practical guidance on how high-level ethics principles can be translated into verifiable criteria" [1, p. 9]. the CapAI procedure consists of three components an internal review protocol (IRP), a summary datasheet (SDS) and an external scorecard (ESC).

2.3.1 Why CapAI

As discussed in Section 2.2, the AI Act proposes extensive requirements for AI systems. Especially High-risk AI systems are expected to conform to a wide range of requirements. The key enforcement mechanism in the AI Act's toolkit is the conformity assessment. This assessment should make market surveillance easier for authorities and ensure that providers adhere to this legislation. However, the AI Act "neither prescribes nor details the form of such conformity assessments" [28, p. 14]. CapAI tries to fill this gap by aiding firms required to conduct AI Act conformity assessments. This is done by proposing a procedure, which involves the entire AI lifecycle, for assessing conformity with the Act and creating the necessary documentation to prove compliance. Figure 2.3 shows which



(a) high-risk AI systems, internal control

(b) high-risk AI systems, external control

(c) Limited and minimal risk AI systems

Figure 2.3: Coverage of CapAI with regards to the AI Act requirements for specific systems. Source: [28, p. 14/16]

requirements for each type of system are tackled by the CapAI procedure. CapAI does not provide the quality management or post-launch monitoring system which will be mandatory under the AI Act. But does address the necessary documentation for these systems. Even though the AI Act does not establish any hard requirements for limited or minimal risk systems, CapAI does recommend that providers of these systems implement the procedure of CapAI as best practice in the form of a voluntary code of conduct. Furthermore, CapAI can be used both for internal and external conformity assessments. While providers could utilize CapAI to shape their self-assessments, notified bodies could use the procedure in the same way. In the case of High-risk AI systems CapAI does not provide a framework for the implementation of quality management systems [1, art. 17] or monitoring/logging systems [1, art. 12] as required by the AI Act.

2.3.2 Internal Review Protocol

The internal review protocol (IRP) follows all development stages of the AI system's life cycle. Figure 2.4 shows the AI lifecycle as defined by CapAI. For each stage of the life cycle, the IRP defines requirements that assess the relevant ethical issues. The protocol "helps organisations to assess the awareness, performance and resources in place to prevent potential failures, as well as the process for responding and rectifying potential failures" [28, p. 16].

The IRP serves as a confidential document that has limited accessibility. However, similar to accounting data, it may be disclosed in a legal context to facilitate business-to-business contractual agreements or as evidence in addressing legal disputes associated with audits of the AI system. This confidentiality means that for every requirement of the IRP, a specific key actor is defined to answer it. The stakeholders set out by CapAI are [28, p. 17]:

- "**Top manager responsible for AI**, who bears responsibility for justifying the application

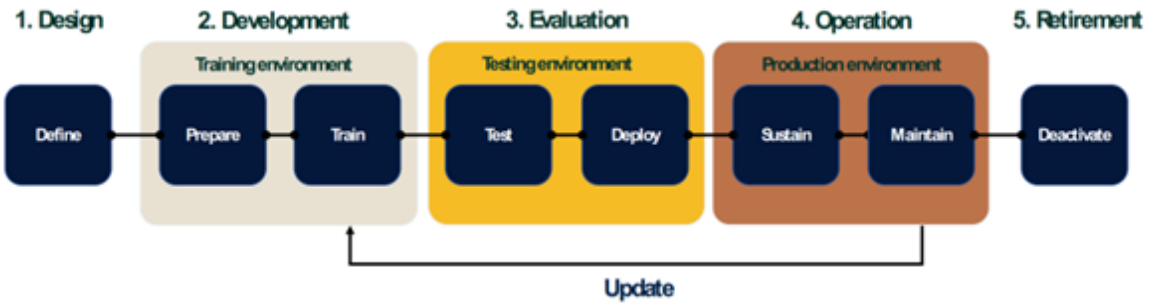


Figure 2.4: The CapAI AI life cycle with its five stages and key steps. Source: [28, p. 17]

and performance of the AI system to all stakeholders, internally and externally."

- "**Product owner**, who is responsible for the performance of the AI system in question."
- "**Project manager**, who leads the development (or, if externally sourced, procurement) process."
- "**Data scientist**, who leads the technical implementation of the AI system in question."

The descriptions of these actors may differ from the regularly accepted definitions. However, it seems that CapAI only uses these key actors as loose contours for which respondents should be involved with the execution of the IRP. Organisations will have to modify the CapAI procedure to adhere to their own needs and use cases.

Each requirement item consists of an item description, the request for supporting information as evidence for the completion of the item and the target respondent which oversees that the requirement in question is met. In practice, the IRP can function as a checklist which can be completed stage by stage chronologically. An overview of all IRP requirement items can be found in Appendix A.

2.3.3 Summary datasheet

The summary datasheet (SDS) is a high-level summary of the AI system's purpose, functionality and performance. The SDS is meant to store all information needed for the registration of high-risk AI systems in accordance with the AI Act [1, Art. 51]. The information which needs to be included in the SDS is derived from the AI Act [1, Annex VIII] itself and outlined by CapAI as the following [28, p. 27]:

- "Name, address and contact details of the provider."
- "Where another person carries out submission of information on behalf of the provider, the name, address and contact details of that person."
- "Name, address and contact details of the authorised representative, where applicable."

- "AI system trade name and any ambiguous reference allowing identification and traceability of the AI system."
- "Description of the intended purpose of the AI system."
- "Status of the AI system (on the market, or in service; not placed on the market/in service, recalled)."
- "Type, number and expiry date of the certificate issued by the notified body and the name of identification number of that notified body (where applicable)."
- "A scanned copy of the certificate referred to in point 7 (where applicable)."
- "Member States in which the AI system is or has been placed on the market, put into service or made available in the Union."
- "A copy of the EU declaration of conformity referred to in Article 48."
- "Electronic instructions for use; this information shall not be provided for high-risk AI systems in the areas of law enforcement and migration, asylum and border control management referred to in Annex III, points 1, 6 and 7."
- "URL for additional information (optional). Providing this link is optional, yet in our view it is useful to include it here as well as in the external scorecard, which we are proposing below as an additional document to be made available publicly."

2.3.4 External scorecard

The external scorecard (ESC) is a document summarising the purpose and the key aspects of the ethical values behind the development of the AI system. The ESC is meant to be made available externally for any relevant stakeholder such as customers or business partners. It functions as a "'health check' to show the application of good practice and conscious management of ethical issues across the AI life cycle" [28, p. 28]. The ESC does not disclose competitive or sensitive information about the AI system of the organisation in question. ESCs are similar to model cards [29] which detail performance characteristics of machine learning models. The elements displayed by the ESC can be chosen for each AI system specifically and according to the underlying circumstances. CapAI does suggest four "meaningful aspects" to be made available using the ESC. These aspects are shown in Figure 2.5.

The CapAI procedure states that the answers to the ESC aspects should be generated through the IRP [28, p. 16]. The ESC summarises the relevant information gathered by the IRP into an overall risk score. Therefore, the ESC should be assembled after completing the IRP [28, p. 17]. By formulating the response on these ESC aspects in an understandable manner catered to the end user of the system, it should bring forth a clear understanding of the use case of the system and the ethical values which shaped its development. Utilising the ESC in this way should create a baseline of transparency for stakeholders to make informed decisions in the usage of the product and prevent them from misusing the system.

Item	Action
1. Purpose	Describe the AI system in terms of its objective and functionality.
2. Values	Outline the organisational values and norms that underpin the development of the AI system.
3. Data	A. Define the data used in terms of its public, proprietary and/or private nature. B. State whether the data used is internal and/or provided by a third party. C. Specify how consent has been secured for the use of this data. D. State whether the AI system uses protected attributes.
4. Governance	A. State the person responsible for the AI system. B. Provide a point of contact for any complaints or concerns. C. State the date when the initial AI system was deployed. D. Specify the dates of the last and next review of the AI system.

Figure 2.5: The suggested aspects of the CapAI ESC. Source: [28, p. 28]

2.3.5 Limitations

CapAI is an Ethics-based Auditing procedure [28, p. 71] [30]. Several risks and limitations have been defined for these types of frameworks, and thus, for CapAI. In this Section, we will discuss these risks and limitations.

Firstly, auditing procedures are dependent on the intent of different stakeholders [28, p. 71]. Misalignment between the goals of stakeholders and ethical principles could lead to problems such as ethics shopping, ethics bluewashing and ethics lobbying [31]. These problems could influence how the assessment is conducted in practice and how strictly auditors look at the documentation in question. These risks could start to become severe when too much pressure is placed on organisations to implement procedures that they do not have the resources or capacity to support [28].

Secondly, the results of an IRP are subject to the potential for adversarial behaviour [28]. By withholding information, changing behaviour or supplying false data during the audit process the outcome of the audit can unjustly turn out more positive for certain stakeholders. Even when this so-called management fraud does not occur, corrective steps may even be prevented by power asymmetries [28].

Thirdly, CapAI primarily focuses on "what" organisations should have to ensure ethical AI. However, there comes a time that researchers will have to focus on "how" organisations should execute these requirements [32]. Only using the IRP as a checklist for necessary documentation without knowledge of what this documentation should entail could lead to paper tigers that lack substance or value. Without a deep understanding of the documentation requirements, organisations risk creating a hollow framework that fails to address the intricacies of their unique risks and vulnerabilities. Therefore, we believe it is imperative to combine the IRP with a thorough

comprehension of the underlying principles and best practices, allowing for the creation of a resilient risk management system that truly safeguards ethical AI.

A clear and crucial example of CapAI not defining "what" certain parts of the procedure should entail is the lack of stakeholder descriptions. CapAI mentions stakeholders multiple times but never explicitly defines what stakeholders are in the context of CapAI. While the procedure does mention both internal and external stakeholders to be relevant for t least some parts of the procedure, it is not clear how CapAI's tools should be utilised to meet the stakeholder's needs. Additionally, the ESC's essential aspects as defined by CapAI, as seen in Figure 2.5, seem to be very limited in their descriptions. The four present aspects possess no reference to crucial principles like robustness, fairness or privacy. Even though these principles are explicitly mentioned in the AI Act to be of necessity.

At last, even though CapAI provides useful tools for the swift adoption of an ethical conformity assessment, the process is likely to require additional resources from companies who want to implement it [3]. Therefore, the question remains how capAI be supported and implemented in a company, specifically by SMEs, and what external support and tools are needed to successfully adopt CapAI [3].

Chapter 3

Design and development

In Section 1 we discussed that CompAIs goal is to simplify the execution of Conformity assessments and to aid the communication around the AI Act within project teams. In this Section, we will discuss the methods used to achieve this. And more specifically the methods used to develop the CapAI Maturity Model and CompAI system.

3.1 CapAI Maturity Model

Various procedures have been defined for the development of such maturity models [33, 34, 35, 36, 37]. However, for this research, the maturity model is not built from the ground up but expands the existing CapAI IRP. Therefore, some of the conventional steps for developing maturity models are already defined within the CapAI IRP. For the development process of the model, we combined different stages and procedures by various authors [33, 34, 35, 36, 37]. Below we will walk through the decision options of these development phases and describe how they shaped the development of the model.

3.1.1 Objectives

One of the objectives of this research is to ease communication between AI project stakeholders. By expanding the CapAI IRP requirements with maturity levels we hope to create more planes for comparison of AI systems and providers. Using only the CapAI IRP describing certain aspects of AI Act compliance becomes binary. The IRP only measures if certain documentation is present. The maturity model provides users with a way to also describe how this documentation is managed. In addition, by applying the model to an AI system it becomes possible to summarise the state of AI Act compliance with the use of the resulting maturity values. This enables project teams to quickly communicate the past, current and target state of AI Act compliance with higher management and other stakeholders, without these stakeholders needing complete comprehension of the legislation.

Moreover, the usage of this maturity model would give project teams a step-by-step guide for AI compliance as well by breaking each requirement down into manageable stages. This should streamline the communication around AI Act compliance and ethical AI by providing a concrete and standardized path.

In Section 2.3.5 we discussed how the CapAI framework only offers a checklist with items that outline “what” AI providers should have. With this maturity model, we hoped to achieve a way to

describe "how" AI providers should manage these requirement items to truly ensure ethical AI governance. Overall, the composition of management documents adds no value to the organisation when the values of these documents are not followed in practice. The maturity model aims to prevent AI providers from focusing on creating these paper tigers without putting them to practice.

3.1.2 Audience, scope and success criteria

The scope and corresponding audience of the model are identical to that of the CapAI IRP[28]: any provider of an AI system needing or wanting to conduct an AI Act self-assessment or a third party tasked with auditing an AI system conform to the AI Act.

This means that users of the maturity model would be managers of AI systems or external auditors. Within the maturity model itself, we decided to step away from the pre-defined respondents of the CapAI IRP. This decision will be justified later in Section 3.2.

The requirements of the maturity model are pre-defined by the CapAI IRP as well. These requirements are illustrated in Appendix A.

3.1.3 Maturity levels

As discussed in Section 3.1.1, the model should have a descriptive, prescriptive and comparative use [38]. Since the base of the model has been laid down in the CapAI IRP these use cases have to be fulfilled in the definition of the maturity levels. To conserve the chronological nature of the IRP we decided to keep the object division of the model into the stages of the AI lifecycle. For the dimensions of the maturity model, we mapped 5 levels to each of the 40 requirement items of the CapAI IRP (as illustrated in Appendix B). The dimensions of maturity are:

1. **Initial:** unstructured approach, no documentation defined;
2. **Repeatable:** an approach has been defined but not formally accepted;
3. **Defined:** the documented approach has formally been accepted and is being used in practice.
4. **Managed and Measurable:** the approach is adopted by the organisation, results are reviewed and updated regularly.
5. **Continuous improvement:** There is continuous improvement in the defined approach.

The maturity model corresponds to a single AI system, which is a rare use case. Primary examples of AI ethics maturity models focus on the organisational level [39, 40, 41]. Therefore, it was not possible to base the level descriptions on existing models. Since the AI Act bears much resemblance with the GDPR [14] we decided to derive the maturity levels from existing GDPR maturity models [42]. The requirement text for each level was derived in the same way from these models. In formulating the requirement it was necessary to keep the three use cases in mind (descriptive, prescriptive and comparative). However, the scope of this research did limit the possibility of adding substantiation to each of the requirement texts.

3.2 CompAI

The goal of creating the CompAI system was to implement the CapAI procedure [28], in a minimum viable product that streamlines communication surrounding the AI Act for project teams. In this context, CompAI would be considered a minimum product when it enables users to work with the 3 CapAI tools, specifically the IRP, SDS and ESC. Additionally, CompAI is considered a viable product when our review process indicates that it possesses high perceived usefulness from industry professionals. In this section, we will discuss what methods were used to achieve this. First, we will explain the technical specifications of the system. Following up with the different features implemented in CompAI.

3.2.1 Technical specifications

The technical design of the CompAI system is fairly simple. CompAI is built as a web app to provide the possibility for project teams to collaborate within the same environment. The back- and front-end are coded using the open-source high-level Python web framework Django [43]. This is done because Django offers a fast workflow to bring applications from concept to production, which was necessary for the time frame of the research. Furthermore, the scalability and popularity of the framework allow organisations to implement their own CompAI versions, conforming to their specific requirements and building on our open-source system.

To speed up the front-end development we utilised the free and open-source web application UI kit called Tabler [44]. This UI kit offers a wide range of components for creating dashboard apps based on Bootstrap 5. This enabled a quick implementation of a responsive UI using prefabricated components and layouts. This kit is used under the MIT license [45], which ensures limited restrictions on reuse.

For the usage of graphs and other visualisations, Apexchart.js [46] was used. This is a JavaScript library for building interactive data visualisations and charts. This library was again used under the MIT licence [45]. Apexcharts.js was the preferable library for this project because of its popularity and the possibility to export the generated charts from within the web apps interface.

For exporting the SDS and ESC as PDF files we made use of an open-source library called ReportLab PDF toolkit [47]. This is a library for creating graphs and paragraphs of Python objects and rendering them to a PDF file. The library operates under an Open Source License, which allows us to use, modify, and distribute the library for both personal and commercial purposes. ReportLab is a library with a steep learning curve, to speed up the implementation of its features we incorporated parts from a GitHub project which offered an example implementation [48]. This GitHub project is used under the MIT [45] license as well.

3.2.2 Features

To implement the CapAI procedure into the CompAI software, four distinct features were necessary. Below we will go through these features and discuss the methods used in their implementation.

Dashboard

When opening the web app the user is first directed to the dashboard page. Here the user will have an overview of their environment, The dashboard is meant to display all relevant information about

the user's AI systems (called projects within CompAI) that are registered in CompAI. Figure F.1 in Appendix F shows stills of this dashboard. Below we will go discuss each component from top to bottom.

First, we have three radial charts respectively displaying the following data, by hovering over the radial chart the data after the colon is shown:

- the total registered projects for this organisation: projects the user is a member in, not a member in, is the creator of;
- The total created IRP assessments: that are completed, of 50% complete, never filled out;
- The average maturity of the organisation: the percentage of projects with a maturity lower than 3, the percentage of projects with a maturity of 5.

Second, a line chart shows the average maturity of every registered project per IRP stage (as discussed in Section 2.3.2. This chart shows the user which projects are registered along with the results of their most recent IRP assessments.

Third, the dashboard shows a column chart which breaks down for every project how it scores on average maturity per IRP stage. This chart essentially displays the same data as the aforementioned line chart. However, it was added to provide the user with a wider range of visualisations to choose from.

Fourth, a table of the total maturity results of the latest IRP Assessment for each project is shown along with a progress bar indicating the percentage of IRP items answered in the assessment.

At last, there is a column chart showing the average maturity per IRP stage. The data for this is accumulated from the latest IRP Assessment of each project. This chart is meant to give the user insight into how they manage the entire AI life cycle, which stages need to be worked on to further the effort to compliance and which stages are up to standard.

By combining these components the user is offered a set of tools for evaluation and communication on the topic of AI Act compliance. These tools could be essential to convey the status of AI Act compliance to management and other stakeholders. Furthermore, the dashboard could help acquire leadership buy-in for compliance initiatives. This is fundamental to provide these initiatives with the necessary resources to be successful [49].

IRP

The IRP assessment feature can be accessed through the projects page. The projects page shows a list of all registered projects along with some basic information and statistics about these projects. This is shown by Figure F.2 in Appendix F. The page features a button to create new projects.

By selecting a project the user is redirected to the detail page of the specific project. Figure F.3 in Appendix F shows this page where the relevant information of the project can be updated and IRP assessments can be created.

When selecting an assessment the user is again redirected, this time to the IRP Assessment page. Figure F.4 in Appendix F shows that at the top of this page, the user can see which part of the AI lifecycle is being assessed currently and can navigate to other stages if necessary. At the left of the page, the item description and deliverable are displayed. These are directly taken from the CapAI IRP as discussed in Section 2.3.2. Below this, some statistics are provided about the current

assessment, the number of items already filled in, the average maturity of the project so far and a radar chart displaying the distribution of the maturity across the AI lifecycle stages.

In the center of the screen, the user will find the Assessment form. This form is designed according to the standard guidelines for web form design [50, 51] Every item of the current stage is displayed one at a time to prevent the user from being overwhelmed [50]. Using pagination the user can navigate to a specific item. the five maturity levels and their requirements for the selected item are displayed vertically. At the righthand side of the level descriptions, the form fields are displayed.

There are three form fields. First, there is one dropdown field for the maturity level, which prevents the user from inputting unwanted entries. Next, there is a large text field, where the user can give a summary of the status of the requirement item. This is done to keep a record of accountability within the IRP. At last, the user is asked for a link to relevant documents which can serve as proof.

We explicitly decided to not provide the service of uploading documents to CompAI directly. This has 2 distinct reasons. First, allowing users to upload documents leads to multiple risks for both the user and the system. Users could upload malicious data, either intentionally or unintentionally. Furthermore, this would make CompAI responsible for keeping record of these important management documents. The second reason that the uploading of files by users is undesirable is that most organizations already have a file management service in place like sharepoint. From a business continuity standpoint, it would not be wise to undermine these structures by having a separate system where these documents are located. This could lead to problems when several versions are introduced of these documents.

Most file management services provided dynamic linking to files which means that upon moving a file the provided link would still function in most cases. Paired with the information provided by the user in the summary field. We believe this should supply enough evidence and foundation for the IRP.

The user has to submit each item after altering the fields by using the save button. Upon successful submission, the righthand banner will turn from red to green to signify that this item has been saved.

Compared to the CapAI IRP we decided to step away from the defined respondents for each item. Most organizations have their own structures and defined roles. CompAI tries to enable users to work according to their own (collaborative) workflow without restricting them to certain complicated procedures. Corporate governance of AI is involves multiple stakeholders [52]. By not enforcing certain respondents we believe fluent collaboration between these stakeholders is encouraged.

SDS

The methods used for the SDS feature are fairly simple. Using the navigation bar at the top of the web app the user can navigate to the SDS overview page which is shown by Figure F.5 in Appendix F. Here the user is presented with two tables. Left, there are the projects which have been registered in CompAI using the projects page. Right, there are SDS templates which can be created on this page by pressing the top-right button. Since generating Summary data sheets will most times require the same information it is possible for users to create certain templates to ease this process.

When selecting either a template or project the user is redirected to the page for filling out the

SDS form which is shown by Figure F.6 in Appendix F. Here the user finds a simple form which asks the user to fill out all information needed for the SDS as defined in Section 2.3.3. When filling out the SDS for a project the user can import an template using the "load template" button. After completing the SDS the user can export the SDS as a PDF using the "export to PDF" button. Appendix D shows an example of an exported SDS.

ESC

The ESC feature is rather similar to the SDS. Figure F.7 in Appendix F shows the overview page which again shows tables of the projects and ESC templates. Figure F.8 in Appendix F shows the page for filling out the information required for an ESC.

As discussed in Section 2.3.4 the ESC should be generated through the IRP by utilising the information provided there. However, we believed this would not provide a friendly user experience. The ESC is meant to be published externally. This would mean that most organisations would prefer to be able to change their tone of voice or ways of explaining the system depending on their audience. Furthermore, as with the CompAI IRP implementation, we wanted to provide users with an open procedure to give freedom to the user to use the system as they pleased. This means that we do not force the user to complete the IRP before generating an ESC. The process for generating the ESC is kept simple and offers the user freedom in its execution.

The elements of the ESC are the same as provided by CapAI and are grouped accordingly to give the user a better overview of the form [51]. As with the SDS the user is able to import templates and export the ESC to a PDF for which an example can be found in Appendix E. The design of the ESC is basic. However, when CompAI is adopted by organisations they could modify the system to generate a PDF according to their own corporate identity.

3.2.3 User workflow

Combining all of the features described in Section 3.2.2 we can compile a user workflow detailing how potential users would work with CompAI. First, upon opening the CompAI web app, the user is greeted, as shown in Figure 3.1, with a login screen in which they put in their credentials. Next, the user sees the empty dashboard, Figure 3.2a, which will fill in with project data when projects are registered.

The user will then navigate to the project page using the navigation bar at the top of the page. Figure 3.2b shows the projects page in the initial status. By pressing the "create project button" the user will be presented with the form as illustrated in Figure 3.3a. Upon creation of the new project, the user is redirected to the project detail page from Figure F.3. Here the user can change the project information and create IRP Assessments by pushing the "Create new Assessment" button. After filling in the Assessment name and selecting the CapAI framework the user will be redirected to the Assessment page from Figure 3.3b. The user will walk through every item and all 5 lifecycle stages to complete the IRP.

After the project is created the user could also choose to create an SDS which is done from the page shown in Figure 3.4a. Here the user can either create a template or decide to generate an SDS. In the last case, the user would see a screen similar to that in Appendix F.6. Alternatively, the user could choose to generate an ESc for the registered system. This would be done from the, at this moment almost empty, ESC overview page illustrated in Figure 3.4b. After generating either the

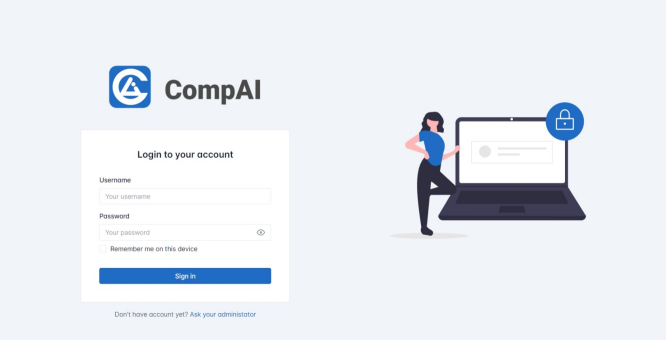
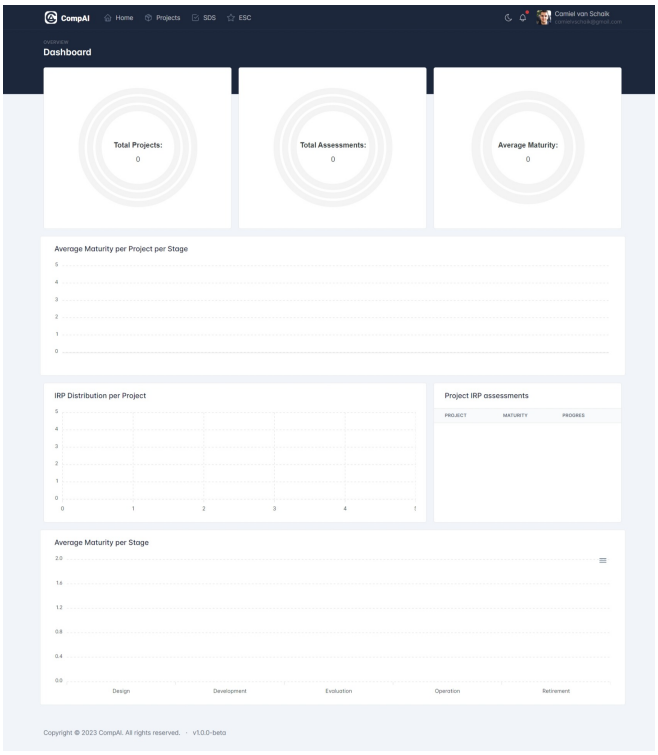


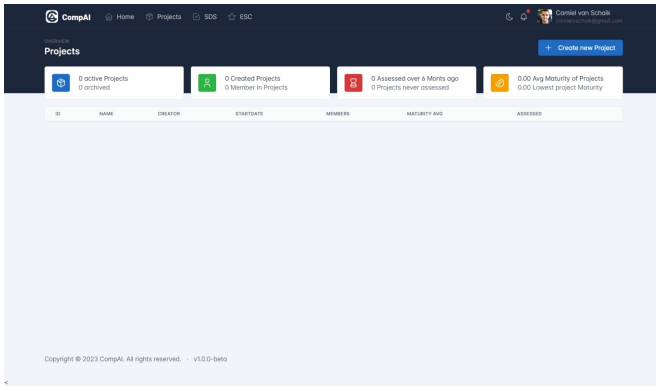
Figure 3.1: The CompAI login screen

SDS or ESC the documents can both be documented in the user's own file management system and be published externally to stakeholders of the system.

With more Projects being registered and IRP being executed the CompAI dashboard will fill with data about these projects. ultimately an active dashboard will start to look like Appendix F.1. The visualisation from both the dashboard and the IRP pages can be used to communicate the status of compliance to higher management and other stakeholders.

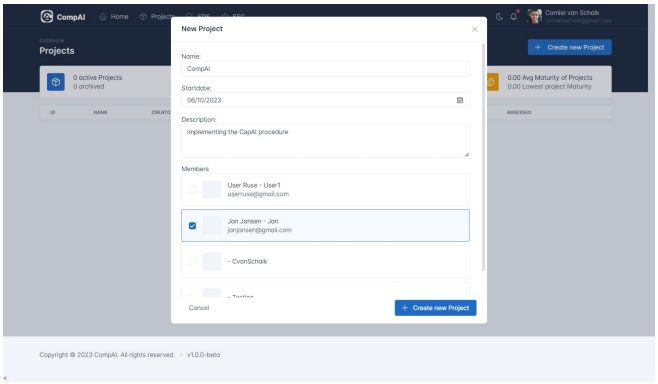


(a) The CompAI dashboard, as seen when no projects have been registered.

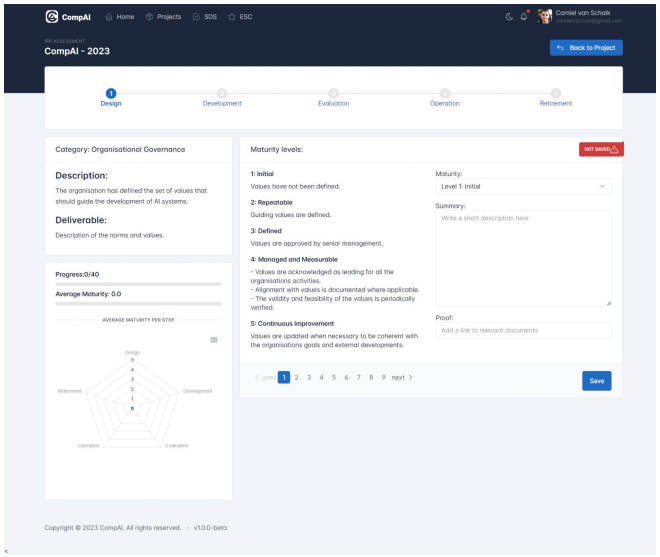


(b) The CompAI projects page, as seen when no projects have been registered.

Figure 3.2

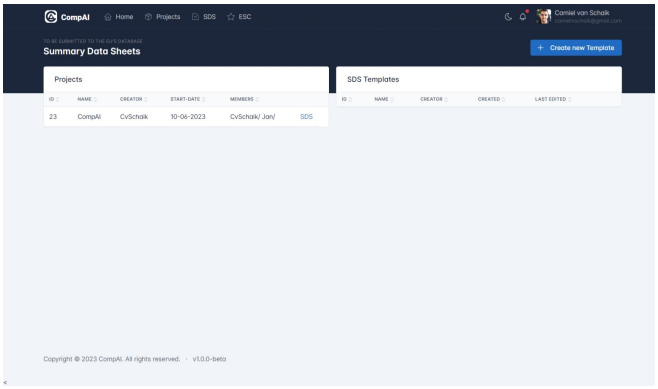


(a) The CompAI project creation form.

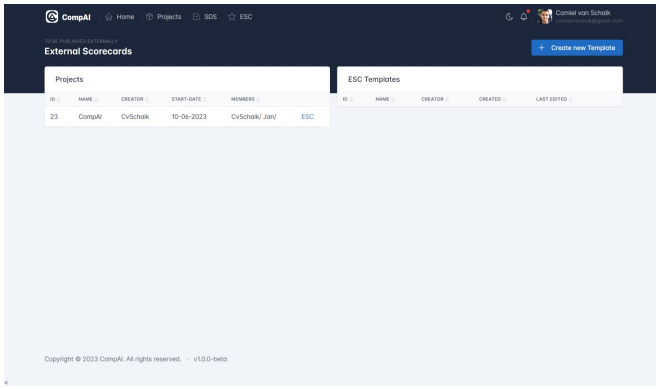


(b) The CompAI IRP Assessment page.

Figure 3.3



(a) The CompAI SDS overview page with only one project registered and no templates made.



(b) The CompAI ESC overview page with only one project registered and no templates made.

Figure 3.4

Chapter 4

Review Results

The last stage of this research is the evaluation stage. This stage is meant to "observe and measure how well the artifact supports a solution to the problem" [4, p. 10]. In Chapter 1 we defined our solution as "a method to simplify AI act compliance by implementing CapAI principles into a streamlined software solution". Two companies were willing to participate in this research and provided feedback on the proposed tools.

Furthermore, these organisations functioned as use cases to apply the accumulated knowledge about the AI Act and the ethics landscape to the real world. The feedback stage consisted of ca. 1.5-hour interview sessions. The first part of the interview session was a presentation about the intricacies of the AI Act (as described in Section 2.2) and how the new regulation will affect the involved company. The second part of the interview involved a demonstration of CompAI with the built-in maturity model. Concluding the session with some time for questions and discussions about the tools as well as a questionnaire.

The effectiveness of this research is measured by the review sessions that were conducted with the two companies. In this chapter, we will discuss the results of those review sessions.

4.1 Respondents

Ultimately two organisations were open to participating in this research and providing feedback during the interview sessions. In this section, we will describe the two companies and what their relation is to the AI Act.

4.1.1 DEUS

The DEUS initiative is an organisation providing ICT-services and consultancy for "human(ity)-centered AI". DEUS assists their clients to utilise data and artificial intelligence for the benefit of people, business and society. To do this DEUS advises organisations across the end-to-end product and service lifecycle using interdisciplinary teams of data scientists, engineers, designers and strategists. The focus of these projects is value creation using data and AI while operating with integrity.

DEUS has locations in The Netherlands, Portugal and Spain and works with both private and public organisations. Their projects cover a wide range of use cases. Some of these projects would be considered High-risk AI systems under the AI Act. For example, their computer vision tool,

Theia, monitors the use of personal protective equipment at industrial mega-sites to reduce safety violations. The system supports safety supervisors by alerting them when possible violations occur. This means that the system could be considered in the categories of **Employment, workers management** or **Management and operation of critical infrastructure** under the AI Act, depending on the actual usage of the system in practice. Projects like Theia make it for DEUS very important to be able to consult their clients on AI Act compliance. CompAI could help their clients with conducting assessments after the project is concluded and the AI systems are deployed. all of this makes DEUS a fitting respondent for our research due to their close relation with responsible AI.

Our respondent for DEUS was the Reliable AI Lead at DEUS which means they primarily lead the research and development of the organisation. Furthermore, they are a PhD Candidate in Responsible AI at the Delft University of Technology, researching the manner in which different types of abductive inferences are generated and evaluated.

4.1.2 BUKO

BUKO is a Dutch company that provides services, facilities and equipment for permanent, temporary and urgent situations. They function as an independent service provider for construction, civil engineering, industry and government, specialised in traffic facilities, transport and camera surveillance.

While BUKO's primary activities do not involve AI, they will be one of many organisations outside of the AI market affected by the regulation. The company consists of 3 branches: BUKO Infrasupport, Transport and Waakt. For the purpose of this research, we primarily focused on the BUKO Infrasupport branch which market leader in modern traffic measures to ensure the safety of road workers and road users. BUKO Infrasupport makes use of a limited amount of Artificial Intelligence. For example, their product called "BUKO Bereikbaar" which tries to reduce hindrances during road works, uses different types of AI and algorithms. Elements of this product are a virtual colleague supported by AI which functions as a chatbot for providing support by answering questions about roadwork projects; a tool which assigns road users automatically the best route given the changing situation around the road works or a system monitoring traffic situations to assess the effectiveness of measures and manipulate traffic flow in real-time.

These automations place BUKO Bereikbaar in the category of limited risk under the AI Act. This means that the organisation would need to adhere to the transparency requirements as discussed in Section 2.2.4. However, looking at the future it would not be unlikely that these systems could be expanded to make invasive decisions and function more independently. For example, if BUKO Bereikbaar would also start managing road or worker safety in an automated way with the use of AI, this would push the product into the category of high-risk AI.

Consequently, BUKO makes a great respondent for our review. The company has started to adopt AI in small steps but does not have it as its primary focus. This means that BUKO has an entirely different position towards the AI Act compared to our other respondent DEUS.

For BUKO my respondent was the Product Manager at BUKO Infrasupport for their product BUKO Bereikbaar.

4.2 Verbal feedback

During the Review sessions, the respondents provided several views and feedback points with regard to the research and CompAI. In this section, we will describe the individual viewpoints that were provided.

Market relevance

Both respondents pointed out that all organisations involved with AI will have to make use of tools like CompAI. With the AI Act coming into effect most organisations affected by the regulation will have to figure out how to achieve compliance. The respondents noted that as with the GDPR 5 years ago, tools like CompAI can speed up this process by taking the organisation through all the necessary requirements. However, one respondent noted that compliance tools such as CompAI could become available in abundance when the AI Act goes into effect. They noted that CompAI may not offer an entirely unique aspect when the market for these tools becomes active.

SME support

During the review sessions, respondents emphasised that CompAI would be especially helpful for SMEs. These companies oftentimes do not have the resources to appoint specific employees to compliance. CompAI could assist in these situations according to the respondents by walking the organisation through all of the requirements of the regulation. Furthermore, the possibility to visualise the compliance status and create an overview of all projects within the organisation was found powerful for communication with management and other stakeholders.

Documentation requirements

Nevertheless, the respondents noted one aspect in particular that could be crucial for organisations working with CompAI. While the possibility to link to proof-documentation within the IRP feature was regarded as very helpful by the respondents, they noted that it would be paramount to incorporate a way to explain to the user what this documentation should entail. One of the respondents indicated that most organisations would lack the knowledge to compose the complete documents that would cover the AI Act requirements. Furthermore, the respondents noted that users of the systems would still need some understanding of the AI Act to be able to decide in which risk class their AI systems would be.

Automation

Furthermore, one respondent suggested that completing an IRP could be a time-consuming task for some organisations. Therefore, they noted that it might be interesting to incorporate generative AI into the system. This way the user could provide the system with a description of the organisation and the project in question. Then, CompAI could pre-fill the IRP and point out potential pitfalls to the user.

Necessity

The respondent from BUKO pointed out that at this time a system like CompAI would not be relevant to their organisation. Due to the AI Act not being in force yet, there would not be value for them to invest in a system like CompAI. The primary reason given for this choice was that AI Act compliance at this time would result in more work and costs while not resulting in more revenue. They referred to the GDPR, where only recently the market has started to en masse demand compliance from suppliers. Only when the AI Act becomes a standard and non-compliance would lead to a loss in market share or severe penalties they believed most companies would not be concerned with using a system like CompAI.

4.3 Questionnaire

The questionnaire gathered feedback from the participants during the review sessions. The goal of the questionnaire is to quantify the feedback about the proposed solutions. This is done using both open and closed questions. The open questions are a simplified and summarised version of the Technology Acceptance Model (TAM model) [53]. This is done to limit the metrics for Perceived Usefulness and Ease of Use both to one statement each, preventing the participants to be overloaded with questions. The summarised TAM model results in the following open questions:

- Using this product would make it easier to do my job.
- It would be easy for me to become agile with the product.

To find out if participants would expect other functionalities within the system, they are also asked to respond to the following question:

- CompAI's capabilities meet my requirements.

These questions are answered using the Likert scale [54] along with the explanatory text. This is done to be able to quantify these answers but also give the participants a method to explain their opinion.

The closed questions are based on the System Usability Scale (SUS) [55] for measuring usability. The SUS model can be used to quantify the usability of CompAI even with a small sample size of respondents [55]. The SUS questions are expanded with three questions that quantify the perceived usability of the four main features of CompAI: the dashboard, the IRP, the SDS and the ESC functionalities. The complete questionnaire can be found in Appendix C. By combining these open and closed questions we hope to quantify how our solutions are perceived by the respondents of the Review sessions.

4.3.1 Results

Here we will showcase the results of this questionnaire. Section 5 interprets these results to assess if CompAI has accomplished the goals of this research.

Closed questions

The questionnaire starts with 14 closed questions The results of the closed questions can be found in Table 4.1.

Using the SUS model on the first 10 questions we can interpret these scores to find the perceived usability. This process sounds somewhat complicated but is fairly simple in practice. Calculating the SUS score starts by converting the answers to a numerical value ranging from 0 to 4 according to the Likert scale [54], with "strongly disagree" being 0 and "strongly agree" being 5. Using these values we can calculate the SUS scores as follows:

$$SUS\ Score = ((X - 5) + (25 - Y)) \times 2.5$$

with:
X = Sum of the points for all odd-numbered questions
Y = Sum of the points for all even-numbered questions

If we do this for both the answers from DEUS and BUKO we see that DEUS scored CompAI with 75 while BUKO gave the system a SUS score of 87.5.

Questions 11 to 14 can then be used to quantify the usability of the separate CompAI features and compare them accordingly. interpreting these scores will be done in Section 5

Open questions

The questionnaire ends with 3 open questions in which the respondents are asked to give more insight into their view on CompAI and it's perceived Usefulness and Ease of Use. The answers to these questions are shown in Tables 4.2 and 4.3. First, the respondents are asked to answer the question using the Likert scale [54]. This is done to be able to compare the responses with each other. At last, the respondents are asked to give an explanation to give nuance to their opinion and explain their answer.

#	SUS ITEM	DEUS Answer	BUKO Answer
1	I think that I would like to use this system frequently.	Somewhat agree	Somewhat disagree
2	I found the system unnecessarily complex.	Somewhat disagree	Strongly disagree
3	I thought the system was easy to use.	Somewhat agree	Strongly agree
4	I think that I would need the support of a technical person to be able to use this system.	Strongly disagree	Strongly disagree
5	I found the various functions in this system were well integrated.	Somewhat agree	Somewhat agree
6	I thought there was too much inconsistency in this system.	Somewhat disagree	Strongly disagree
7	I would imagine that most people would learn to use this system very quickly.	Somewhat agree	Strongly agree
8	I found the system very cumbersome to use.	Somewhat disagree	Strongly disagree
9	I felt very confident using the system.	Neutral	Somewhat agree
10	I needed to learn a lot of things before I could get going with this system.	Somewhat disagree	Strongly disagree
11	I found the home dashboard and graphs to be usefull	Somewhat agree	Strongly agree
12	I found the home dashboard and graphs to be usefull	Somewhat agree	Strongly agree
13	I found the Summary Data Sheet for the EU database module to be useful	Somewhat agree	Strongly agree
14	I found the External scorecard module to be useful	Somewhat agree	Strongly agree

Table 4.1: Results of the closed questions from the Review Questionnaire. These questions are based on the SUS model [55]. Along with the 14 questions, the answers of both respondents are displayed according to the Likert scale [54].

#	Question	Answer	Explanation
1	Using this product would make it easier to do my job.	Somewhat agree	It'll make compliance easier.
2	It would be easy for me to become agile with the product.	Somewhat disagree	It has very little to do with agility
3	CompAI's capabilities meet my requirements.	Neutral	A legal professional should check that.

Table 4.2: Results of the open questions from the Review Questionnaire with DEUS. These questions are based on the TAM model [53]. The results should give insight into CompAI's perceived Usefulness and Ease of Use.

#	Question	Answer	Explanation
1	Using this product would make it easier to do my job.	Neutral	Depends on the context. I think it is useful for our organisation when it becomes required by law. However, I do think that working with a tool like this would be more suitable for someone else in our organization.
2	It would be easy for me to become agile with the product.	Strongly agree	Quick way to gain insights in our performance regarding the AI act
3	CompAI's capabilities meet my requirements.	Strongly agree	Within an organization as ours, it would be 'required by law' to provide these insights. That is basically our basis for the input we need to provide and the insights we need to have. So yes, the tool meets our requirements (which are provided by law)

Table 4.3: Results of the open questions from the Review Questionnaire with BUKO. These questions are based on the TAM model model [53]. The results should give insight into CompAI's perceived Usefulness and Ease of Use.

Chapter 5

Discussion and Future Research

Taking the results of the review sessions as described in Section 4 into account we will try to define if CompAI has accomplished the goal of offering a way to execute the CapAI procedure in a user-friendly way.

First, we can look at the quantifiable results from the questionnaires. In Section 4.3.1 we have seen that the SUS results of the DEUS and BUKO questionnaires were respectively 75 and 87.5. If we look up these values in Table 5.1 we will find that this means that CompAI scores between an A and a B according to our respondents. This means that the respondents rank our system to be between good to excellent usability.

Furthermore questions 11 to 14 of the questionnaire showed that the respondents found each of CompAI's features to be equally useful. This means that CompAI has succeeded in the quantitative part of the review. However, to truly measure if CompAI fulfils our research goals we should also take a look at the qualitative part of the review. By combining both the open questions with the verbal feedback aggregated during the review sessions we can summarise the following qualitative feedback.

1. While CompAI is viewed as being a useful and relevant tool it does not possess a unique market position.
2. CompAI would be particularly useful for SMEs
3. CompAI still requires knowledge of the AI Act to implement it in an organisation.
4. CompAI could benefit from automated assessments.
5. CompAI will not be used by certain companies when they are not forced to comply with regulations or market standards.

These points provide us with some valuable directions for future research. CompAI could improve its user-friendliness by eliminating the need for AI Act knowledge almost entirely. This could be done by implementing an AI classification module which takes users through the process of assessing the risk classification of an AI system. Moreover, CompAI should give users guidance when developing the documentation that is assessed in the IRP. This means that research has to be done as to what this documentation should entail under the AI Act. in addition, we believe that this would provide CompAI with its unique position in the market. However, the goal of this

research was not to create a unique product but to show that it is possible to implement the CapAI procedure into a software solution to simplify the execution of conformity assessments. Furthermore, we believe that with the current state of AI and automation, it would be unwise to make use of generative AI to automate the IRP assessments or to provide users with assistance executing the IRP assessment or composing documentation. Both the IRP assessment and the formulation of documentation should be a conscious process, executed by the responsible human being in our opinion. At last, the aforementioned point 5 states that a lot of companies are likely to neglect AI ethics when they are not forced to comply with regulations or market standards. We acknowledge that this is likely to happen and CompAI will not change this initially. However, as we have seen with the GDPR, with time the market will start to see these ethical principles as standard practice and demand suppliers to follow them. Therefore, systems like CompAI could be used to promote and ease the practice of ethical AI, speeding up this societal process.

However, organisations that would want to adopt CompAI into their processes would require some type of documentation. For the purpose of this research, we have decided to let this thesis function as documentation and not draft up any external documentation. We believe that by utilising well-known frameworks within our system and dissecting, the AI Act, CapAI and CompAI within this thesis it should provide potential users of CompAI with enough aids to adopt CompAI into their organisation.

Admittedly, our process for measuring the usefulness of the system has not been foolproof. While the SUS Model used in the closed questions provided useful quantitative feedback, the open questions proved to generate less insightful results. We found that even though it was beneficial to shorten the questionnaire to prevent the respondents from being overloaded with questions, the formulation of the open questions could be improved. For example, question 2 from Tables 4.2 and 4.3 proved to be unclear in its formulation to gather feedback about how fast the user would be able to work with the system. Future research would benefit from better-defined open questions by either implementing the original TAM model [53] into the questionnaire and independently asking open questions about the system's perceived usefulness and ease of use. or defining questions that summarise the TAM model in a more accurate way.

Furthermore, our review process regrettably consisted of only 2 respondents. Even though these respondents provided a lot of welcome feedback it would be beneficial for a complete assessment of CompAI's usefulness to work with a larger sample size. Our current sample of respondents had different positions toward the AI Act which meant different use cases of CompAI. However, a larger sample size would provide our research with more feedback data and viewpoints. In addition, the SUS model takes an average of 5 respondents [55]. Therefore both the quantitative and qualitative review would benefit from more respondents.

That being said, the respondents who did participate in our review are industry professionals who do have the credibility to talk on behalf of their respective industries. Therefore we believe that despite this smaller sample size, we can conclude from the aggregated data that CompAI has succeeded in the goal of implementing CapAI in a user-friendly way.

SUS Score	Grade	Adjective Rating
> 80.3	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
< 51	F	Awful

Table 5.1: Interpretation table for SUS scores [55].

Chapter 6

Conclusion

To conclude, our research has shown how conformity assessments can be conducted in a user-friendly way by implementing the CapAI principle. CompAI offers users a clear path for executing the IRP conformity assessments in line with the AI Act. CompAI is unique as a system that implements the CapAI procedure. It shows the user what documentation is needed for compliance with the AI Act in a clear and expeditious manner. The system allows for the generation of an SDS to provide all necessary information to the EU AI Database. In addition, CompAI generates an ESC to visualise an AI system's key elements to relevant stakeholders and users. At last, CompAI aids in the communication around the AI Act by providing a clear overview and visualisation of the compliance status.

CompAI proves that it does not have to be overly complicated to implement CapAI. However, there is much room for improvement before organisations could realise the full potential of CapAI by using our system. Users would still need some understanding of the AI Act to utilise CompAI and need even more understanding of ethics principles to achieve compliance. This is partly because CapAI focuses on the conformity assessment of the AI Act and does not account for other requirements of the regulation. Nevertheless, CompAI could be expanded with features to aid users in the risk classification of their systems and the composition of the necessary documentation under the AI Act. Furthermore, to assess the full impact of CompAI on an organisation and measure its usefulness a review should be done with the use of a bigger sample size.

As of writing this thesis, CapAI has been the leading procedure for conducting conformity assessments in line with the AI Act on the market. Nevertheless, our research has pointed out some points of improvement for CapAI to make its procedure more in line with both the AI Act and the AI market. For example, by explicitly adopting ethics principles into the IRP and ESC users of CapAI would be guided towards the significant risks and attention points within their systems. CompAI could therefore either be updated with a future procedure or be expanded upon by modifying the CapAI procedure within our software.

All things considered, our research has achieved the following goals:

- Dissection of the AI Act with all the requirements for providers of all categories of AI;
- Examination of the CapAI framework for conducting conformity assessments, including an overview of all its available methods.
- Maturity model to facilitate better comparison and communication when it comes to AI Act compliance of systems;

- The open-source CompAI web application, which enables users to utilise the CapAI procedure in an accessible fashion;

The CompAI source code on GitHub can be found using the following link:
https://github.com/COvSchaik/CompAI.git

Bibliography

[1] The European Commission. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

[2] The European Parliament, the Council, and the Commission. *Charter of Fundamental Rights of the European Union*. 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN.

[3] "Complying with future AI regulations in Europe: capAI". In: *CMS Law-Now* (2022). URL: https://cms-lawnow.com/en/ealerts/2022/07/complying-with-future-ai-regulations-in-europe-capai.

[4] Ken Peffers et al. "The design science research process: Amodel for producing and presenting information systems research". In: *First International Conference on Design Science Research in Information Systems and Technology*. 2006, pp. 83–16.

[5] Council of the EU. *Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights*. 2022. URL: https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/.

[6] Nestor Maslej et al. "The AI Index 2023 Annual Report". In: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. 2023.

[7] Fortune Business Insights. *Artificial Intelligence Market Report*. 2023. URL: https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114.

[8] U.S.-EU Trade and Technology Council. *Inaugural Joint Statement*. 2021. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/29/u-s-eu-trade-and-technology-council-inaugural-joint-statement/.

[9] Chang Ma. "Self-regulation versus government regulation: an externality view". In: *Journal of Regulatory Economics* 58.2-3 (2020), pp. 166–183.

[10] Jacqui Ayling and Adriane Chapman. "Putting AI ethics to work: are the tools fit for purpose?" In: *AI and Ethics* 2.3 (2022), pp. 405–429.

[11] Anna Jobin, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines". In: *Nature Machine Intelligence* 1.9 (2019), pp. 389–399.

[12] Lee Rainie, Janna Anderson, and Emily Vogels. "Experts doubt ethical AI design will be broadly adopted as the norm within the next decade". In: *Pew Research Center* (2021).

[13] Peter Cihon et al. "AI certification: Advancing ethical practice by reducing information asymmetries". In: *IEEE Transactions on Technology and Society* 2.4 (2021), pp. 200–209.

[14] Council of the European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679.

[15] Marietje Schaake. "Stanford HAI Policy Brief: The European Commission's Artificial Intelligence Act". In: Standford University Human-Centered Artificial Intelligence (BHAI), Standford, Canada, 2021. URL: https://hai.stanford.edu/issue-brief-european-commissions-artificial-intelligence-act.

[16] Anu Bradford. *The Brussels effect: How the European Union rules the world*. Oxford University Press, USA, 2020.

[17] Mauritz Kop. "EU artificial intelligence act: the European approach to AI". In: Transatlantic Antitrust and IPR Developments. 2021.

[18] Lilian Edwards. "The EU AI Act proposal". In: Ada Lovelace Institute. 2022. URL: https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/.

[19] The European Commission. *New legislative framework*. 2008. URL: https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en.

[20] Natali Helberger and Nicholas Diakopoulos. "ChatGPT and the AI Act". In: *Internet Policy Review* 12.1 (2023).

[21] Natascha Gerlach. "The case of the EU AI act: Why we need to return to a risk-based approach". In: International Association of Privacy Professionals. 2023. URL: https://iapp.org/news/a/the-case-of-the-eu-ai-act-why-we-need-to-return-to-a-risk-based-approach/.

[22] Jennifer Cobbe and Jatinder Singh. "Artificial intelligence as a service: Legal responsibilities, liabilities, and policy challenges". In: *Computer Law & Security Review* 42 (2021).

[23] Amanda Lawson. *EU AI act explained*. 2023. URL: https://www.responsible.ai/post/eu-ai-act-explained.

[24] Andreas Liebl and Till Klein. "Analysis of the impact of the EU AI Act on start-ups in Europe". In: appliedAI. 2022. URL: https://www.appliedai.de/en/hub-en/ai-act-impact-survey.

[25] 2020 Commission work programme. "Impact Assessment of the Regulation on Artificial intelligence". In: European Commission. 2021. URL: https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence.

[26] Benjamin Mueller. *How Much Will the Artificial Intelligence Act Cost Europe?* Tech. rep. Information Technology and Innovation Foundation, 2021.

[27] "European Artificial Intelligence Act: Many procedural and substantive requirements". In: PwC. 2022. URL: https://www.pwc.nl/en/insights-and-publications/themes/digitalization/european-artificial-intelligence-act-many-procedural-and-substantive-requirements.html.

[28] Luciano Floridi et al. "CapAI-A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act". In: (2022).

[29] Margaret Mitchell et al. "Model cards for model reporting". In: *Proceedings of the conference on fairness, accountability, and transparency*. 2019, pp. 220–229.

[30] Jakob Mökander and Luciano Floridi. "Ethics-based auditing to develop trustworthy AI". In: *Minds and Machines* 31.2 (2021), pp. 323–327.

[31] Luciano Floridi. "Translating principles into practices of digital ethics: Five risks of being unethical". In: *Ethics, Governance, and Policies in Artificial Intelligence* (2021), pp. 81–90.

[32] Jessica Morley et al. "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices". In: *Science and engineering ethics* 26.4 (2020), pp. 2141–2168.

[33] Tonia De Bruin et al. "Understanding the main phases of developing a maturity assessment model". In: *Australasian Conference on Information Systems (ACIS)*. Australasian Chapter of the Association for Information Systems. 2005, pp. 8–19.

[34] Tobias Mettler. "Maturity assessment models: a design science research approach". In: *International Journal of Society Systems Science* 3.1-2 (2011), pp. 81–98.

[35] Jörg Becker, Ralf Knackstedt, and Jens Pöppelbuß. "Developing maturity models for IT management: A procedure model and its application". In: *Business & Information Systems Engineering* 1 (2009), pp. 213–222.

[36] Anja M Maier, James Moultrie, and P John Clarkson. "Assessing organizational capabilities: reviewing and guiding the development of maturity grids". In: *IEEE transactions on engineering management* 59.1 (2011), pp. 138–159.

[37] Marlies van Steenbergen et al. "The design of focus area maturity models". In: *Global Perspectives on Design Science Research: 5th International Conference, DESRIST 2010, St. Gallen, Switzerland, June 4-5, 2010. Proceedings. 5*. Springer. 2010, pp. 317–332.

[38] Jens Pöppelbuß and Maximilian Röglinger. "What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management". In: (2011).

[39] J Krijger et al. "The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations". In: *AI and Ethics* (2022), pp. 1–13.

[40] Open Data Institute. *Data Ethics Maturity Model: benchmarking your approach to data ethics*. 2022. URL: https://theodi.org/article/data-ethics-maturity-model-benchmarking-your-approach-to-data-ethics/.

[41] Kathy Baxter. *AI Ethics Maturity Model*. 2022. URL: https://www.salesforceairesearch.com/static/ethics/EthicalAIMaturityModel.pdf.

[42] Tamás Laposa and Gáspár Frivaldszky. "Data Protection Maturity: an analysis of methodological tools and frameworks". In: *Central and Eastern European eDem and eGov Days*. 2020, pp. 135–147.

[43] The Django Software Foundation. *Django*. 2023. URL: https://www.djangoproject.com/.

[44] Paweł Kuna. *Tabler*. 2023. URL: https://tabler.io/.

[45] Massachusetts Institute of Technology. *MIT license*. 2023. URL: https://mit-license.org/.

[46] *ApexCharts.js*. 2020. URL: https://apexcharts.com/.

[47] ReportLab Europe Ltd. *reportlab.com*. URL: https://www.reportlab.com/.

[48] Jurasec. *python-reportlab-example: PDF report example with a front-page, headers and table*. URL: https://github.com/jurasec/python-reportlab-example.

[49] Abhishek Gupta. *How to build an AI ethics team at your organization?* 2021. URL: https://towardsdatascience.com/how-to-build-an-ai-ethics-team-at-your-organization-373823b03293.

[50] Javier A Bargas-Avila et al. "Simple but crucial user interfaces in the world wide web: introducing 20 guidelines for usable web form design". In: *User interfaces*. IntechOpen, 2010.

[51] Mirjam Seckler et al. "Designing usable web forms: empirical evaluation of web form improvement guidelines". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 1275–1284.

[52] Peter Cihon, Jonas Schuett, and Seth D Baum. "Corporate governance of artificial intelligence in the public interest". In: *Information* 12.7 (2021), p. 275.

[53] Fred D Davis. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". In: *MIS quarterly* (1989), pp. 319–340.

[54] Rensis Likert. "A technique for the measurement of attitudes." In: *Archives of psychology* (1932).

[55] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* (1996), pp. 4–7.

Appendix A

CapAI IRP

In this Appendix we included the CapAI Internal Review Protocol (IRP) The IRP consists of 5 stages which represent the entire AI lifecycle from design to retirement. The IRP has 40 distinct requirements. For each requirement, an item description, support and respondent have been defined. The item description tells the user what the organisation should do to be compliant with the AI Act. The support suggests which support documents should be present for this requirement and the respondent is the person who should be responsible for ensuring the requirement is met and documented.

Stage	#	Item	Support	Respondent
Stage 1: Design	1	The organisation has defined the set of values that should guide the development of AI systems	Description of the norms and values	Top manager responsible for AI
	2	These values have been published/ communicated externally	Short description of how values were communicated externally	Top manager responsible for AI
	3	These values have been communicated to internal AI project stakeholders	Short description of how values were communicated internally	Top manager responsible for AI
	4	A governance framework for AI projects has been defined	Short description of the AI governance framework, i.e., how adherence to the organisational values will be ensured and demonstrated in practice	Top manager responsible for AI
	5	The responsibility for ensuring and demonstrating that AI systems adhere to defined organisational values has been assigned	Name(s) of the person assigned	Top manager responsible for AI
	6	The objectives of the AI application have been defined and documented	Short description of the objectives of the AI application	Project manager
	7	The AI application has been assessed against the ethical values	Ethical assessment	Project manager
	8	Performance criteria for the AI application have been defined	Requirement specification document	Project manager
	9	The overall environmental impact for this AI application has been assessed	Assessment of the environmental impact of the AI application	Project manager
pment	10	The data used to develop the AI application has been documented	List of data used in the AI application	Project manager
	11	Data used in the development has been checked for representativeness, relevance, accuracy, traceability (e.g., external data) and completeness	Data impact assessment; see e.g., IAF Ethical Data Impact Assessment or CNIL Privacy Impact Assessment	Project manager
	12	The risks identified in the data impact assessment have been considered and addressed	Handling missing data; handling imbalance data; scaling; normalisation	Project manager
	13	Legal compliance with respect to data protection has been assessed, e.g., GDPR	Data compliance assessment, including a list of protected attributes	Project manager

Stage 2: Develo	14	The source of the model has been documented	Source of the model	Project manager
	15	The selection of the model has been assessed with regard to fairness, explainability and robustness	List of risks identified	Project manager
	16	The risks identified in the model have been considered and addressed	List of assurance countermeasures	Project manager
	17	The strategy for validating the model has been defined	Brief description of the validation strategy	Project manager
	18	The organisation documented the AI performance in the training environment	Performance on the training set in relation to agreed objectives	Data scientist
	19	The setting of hyperparameters has been documented	Justification for the selection and levels of hyperparameters used	Data scientist
	20	The model fulfils the established performance criteria levels	Documentation of model performance	Project manager
Stage 3: Evaluation	21	The strategy for testing the model has been defined	Short description of the validation strategy	Project manager
	22	The organisation has documented the AI performance in the testing environment	Documentation model performance on the testing set in statistical terms	Data scientist
	23	The model has been tested for performance on extreme values and protected attributes	Short description of performance on extreme values and protected attributes	Data scientist
	24	Patterns of failure have been identified	FMEA, e.g., error curves, overfitting analysis, exploration of incorrect predictions	Data scientist
	25	Key failure modes have been addressed	Short description of how to resolve or account for key failure modes	Data scientist
	26	The model fulfils the established performance criteria levels	Documentation of model performance	Project manager
	27	The deployment strategy has been documented	Short description of the deployment strategy	Product owner
	28	The serving strategy has been documented	Short description of the serving strategy	Product owner
	29	The risks associated with the given serving and deployment strategies have been identified	Short description of identified risks	Product owner
	30	The risks associated with the given serving and deployment strategies have been addressed	Short description of how to resolve or account for key risks	Product owner

	31	The model fulfils the established performance criteria levels in the production environment	Performance in the production environment	Product owner
Stage 4: Operation	32	The risks associated with changing data quality and potential data drift have been identified	A short description of the risks associated with data quality is captured (e.g., data drift, bias drift, feature attribution drift)	Product owner
	33	The risks associated with model decay have been identified	A short description of the risks associated with model decay is captured	Product owner
	34	The strategy for monitoring and addressing risks associated with data quality and drift; and model decay has been defined	Outline of monitoring strategy (e.g., error classification, critical threshold values for data drift and model decay)	Product owner
	35	Periodic reviews of the AI applications with regard to the ethical values have been set	Review schedule and format	Top manager
	36	The organisation has a strategy for how to update the AI application continuously	Frequency of updates and documentation of model changes	Product owner
	37	A complaints process has been established for users of the AI system to raise concerns or suggest improvements	Short description of the complaints process (e.g., point of contact)	Product owner
	38	A problem-to-resolution process has been defined	Outline of problem-toresolution process	Product owner
Stage 5: Retirement	39	The risks of decommissioning the AI system have been assessed	Documentation of decommissioning risks	Product owner
	40	The strategy for addressing risks associated with decommissioning the AI system	Outline of the strategy to manage the risks of decommissioning AI (e.g., data residuals: what will happen to data records, model accessibility and interfaces to other systems)	Top manager

Appendix B

Maturity Model

Appendix B shows our maturity model. This model maps the CapAI IRP to our own designed five maturity levels:

1. **Initial:** unstructured approach, no documentation defined;
2. **Repeatable:** an approach has been defined but not formally accepted;
3. **Defined:** the documented approach has formally been accepted and in practice.
4. **Managed and Measurable:** the approach is adopted by the organisation, results are reviewed and updated regularly.
5. **Continuous improvement:** There is continuous improvement in the defined approach.

The maturity model should give users another plane of comparison when evaluating systems of AI providers in terms of AI Act compliance. This should help with communicating the status of AI compliance for a system to relevant stakeholders. Below, each stage of the IRP is divided into colour-coded subdomains. For readability the header of the table is repeated on every page, some domains do span multiple pages.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Organisational Governance	1	The organisation has defined the set of values that should guide the development of AI systems	Description of the norms and values	Top manager responsible for AI	Values have not been defined.	Guiding values are defined.	Values are approved by senior management.	- Values are acknowledged as leading for all the organisations activities. - Alignment with values is documented where applicable. - The validity and feasibility of the values is periodically verified.	Values are updated when necessary to be coherent with the organisations goals and external developments.
Organisational Governance	2	These values have been published/ communicated externally	Short description of how values were communicated externally	Top manager responsible for AI	Values have not been communicated externally.	Values have been communicated externally.	Values are freely available as hard copies and accessible for any external party.	Reflection of these values in procedures and policies are comunicated externally.	Changes and updates of values are communicated externally in a transparent manner.
Organisational Governance	3	These values have been communicated to internal AI project stakeholders	Short description of how values were communicated internally	Top manager responsible for AI	Values have not been communicated internally.	Values have been communicated internally.	Values are freely available as hard copies and accessible for the entire organisation.	The values are part of an internal ethics awareness program.	Changes and updates of values are communicated internally in a transparent manner.
Organisational Governance	4	A governance framework for AI projects has been defined	Short description of the AI governance framework, i.e., how adherence to the organisational values will be ensured and demonstrated in practice	Top manager responsible for AI	There is no governance framework.	A governance framework has been defined.	- The governance framework has been aproved by senior management. - Effectiveness and usage of the. framework is assessed on ad hoc basis.	- The framework is adopted organisationwide and usage is documented. -The framework is periodically evaluated updated and reapproved by management.	- Reports about effectiveness of the framework are submitted to senior management.
Organisational Governance	5	The responsibility for ensuring and demonstrating that AI systems adhere to defined organisational values has been assigned	Name(s) of the person assigned	Top manager responsible for AI	No ownership, roles and responsibility are assigned.	Ownership, roles and responsibility are assigned informally and ad hoc during projects.	Ownership, roles and responsibility are formally assigned both on project and organisational level.	Ownership, roles and responsibility are periodically evaluated.	Senior management reflects on responsibility evaluation.
Use Case	6	The objectives of the AI application have been defined and documented	Short description of the objectives of the AI application	Project manager	No objectives have been defined.	Objectives have been defined but not formally documented.	Objectives have been defined and are formally documented.	Periodic assessments are in place to asses the relevance of the objectives and prevent function creep.	The objectives are updated with the growth of the application or change of scope these changes and their motivation are documented.
Use Case	7	The AI application has been assessed against the ethical values	Ethical assessment	Project manager	There has been no ethical assessment.	An ethical assessment has been executed.	Ethical assessments are executed when necessary to keep up with application changes.	Results of assessments are periodically evaluated and communicated externally.	Evaluations of assessment methods are periodically executed and documented.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Use Case	8	Performance criteria for the AI application have been defined	Requirement specification document	Project manager	Performance criteria have not been defined.	Performance criteria have been defined.	The application is periodically tested against the criteria and results are formally documented.	The performance criteria themselves are periodically evaluated.	The performance criteria are updated with the growth of the application or change of scope these changes and their motivation are documented.
Use Case	9	The overall environmental impact for this AI application has been assessed	Assessment of the environmental impact of the AI application	Project manager	There has been no environmental impast assessment.	An environmental impact assessment has been executed.	Environmetal impact assessments are executed when necessary to keep up with application changes.	Results of assessments are periodically evaluated and communicated externally.	Evaluations of assessment methods are periodically executed and documented.
Data	10	The data used to develop the AI application has been documented	List of data used in the AI application	Project manag	Data used in the AI application has not been documented.	Data used in the AI application has been documented informally and is incomplete or inaccurate.	Data used in the AI application has been documented and is complete and accurate. - documantation includes, the data used, the category of data, the source and reason for usage.	The documentation is periodically reviewed and updated when necessary.	The periodic documantation proces of used data is incorporated into the development cycle.
Data	11	Data used in the development has been checked for representativeness, relevance, accuracy, traceability (e.g., external data) and completeness	Data impact assessment; see e.g., IAF Ethical Data Impact Assessment or CNIL Privacy Impact Assessment	Project manag	No data impact assessments have been conducted.	A data impact assessment has been carried out, but it is incomplete or inaccurate.	A data impact assessment has been carried out, and is complete and accurate.	New assessments are conducted when datasets, used data, sources are changed or added.	Assessment procedures are periodically reviewed to assure completeness, relevance and unbiasedness.
Data	12	The risks identified in the data impact assessment have been considered and addressed	Handling missing data; handling imbalance data; scaling; normalisation	Project manag	No risks have been identified or addressed.	Risks have been identified but not yet addressed.	Risks have been identified and addressed to some extent but there are no mitigations in place.	Risks are identified and addressed in a systematic manner.	The organisation has established a continuous improvement process to identify and address new risks and evolving threats.
Data	13	Legal compliance with respect to data protection has been assessed, e.g., GDPR	Data compliance assessment, including a list of protected attributes	Project manag	No assessment of data protection regulations has been conducted.	Assessment of data protection regulations has been conducted but no compliance measures have	Compliance measures have been implemented to address identified data protection issues.	Compliance measures have been fully integrated into the organization's data handling practices.	Compliance measures are continuously monitored and updated to reflect changes in data protection regulations.
Model	14	The source of the model has been documented	Source of the model	Project manag	There is no documentation of the source of the model.	Ocumentation of the source of the model exists but is incomplete or inaccurate.	Documentation of the source of the model is complete and accurate.	Documentation of the source of the model is integrated into the organization's knowledge management systems.	Documentation of the source of the model is actively maintained and updated to reflect changes in the model's development.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Model	15	The selection of the model has been assessed with regard to fairness, explainability and robustness	List of risks identified	Project manage	No assessment of the model's fairness, explainability, and robustness has been conducted.	Assessment of the model's fairness, explainability, and robustness has been conducted but is incomplete or inaccurate.	Assessment of the model's fairness, explainability, and robustness has been conducted and is complete and accurate.	Fairness, explainability, and robustness assessments are fully integrated into the organization's development practices. - Assessment methods are periodically reviewed.	Fairness, explainability, and robustness are continuously monitored and assessed to keep up with changes in the model's development and external changes.
Model	16	The risks identified in the model have been considered and addressed	List of assurance countermeasures	Project manage	No consideration or addressing of identified risks.	Identified risks have been considered but not fully addressed.	Identified risks have been addressed in the model.	Risks mitigation fully integrated into the organization's model development practices.	Risks are continuously monitored, updated and mitigated to reflect changes in the model's development.
Model	17	The strategy for validating the model has been defined	Brief description of the validation strategy	Project manage	There is no validation strategy defined.	A validation strategy is defined on ad-hoc basis but not documented.	The validation strategy is documented and communicated to stakeholders.	he validation strategy is consistently followed, and its effectiveness is monitored.	The validation strategy is continuously improved based on feedback and experience.
Model	18	The organisation documented the AI performance in the training environment	Performance on the training set in relation to agreed objectives	Data scientist	AI performance is not documented.	AI performance is documented on ad-hoc basis.	AI performance is regularly documented and monitored.	AI performance documentation is integrated into the development process and is used to inform decisions.	AI performance documentation is used to continuously improve the AI system.
Model	19	The setting of hyperparameters has been documented	Justification for the selection and levels of hyperparameters used	Data scientist	Hyperparameters are not documented.	Documentation of hyperparameters used, but not the justification for their selection.	Justification for the selection of hyperparameters is documented.	Selection of hyperparameters is based on established best practices and guidelines.	- Hyperparameter selection is regularly reviewed and updated based on feedback from system performance and user needs. - The selection process for hyperparameters is documented and publicly available for transparency and accountability.
Model	20	The model fulfils the established performance criteria levels	Documentation of model performance	Project manage	Model performance is not documented.	Model performance is documented on ad-hoc basis.	Model performance is regularly documented.	Performance criteria levels are actively monitored, and deviations are addressed promptly. Performance metrics are reported regularly.	- The model performance does not dictate the systems criteria levels. - There are procedures in place to prevent this from happening.
Test	21	The strategy for testing the model has been defined	Short description of the validation strategy	Project manager	No testing strategy is defined.	Testing is done on an ad-hoc basis.	Testing strategy is formaly documented.	The testing strategy is regularly reviewed and updated.	The testing strategy includes ethical considerations, such as testing for fairness and avoiding harm to vulnerable groups.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Test	22	The organisation has documented the AI performance in the testing environment	Documentation model performance on the testing set in statistical terms	Data scientist	No documentation of AI performance in the testing environment.	AI performance in the testing environment is documented but not consistently.	The organisation has a consistent and comprehensive method for documenting AI performance in the testing environment.	The documentation of AI performance in the testing environment is regularly reviewed and updated, and the review process is well-defined and executed.	The organisation has an automated system for continuously monitoring and documenting AI performance in the testing environment, and the system is regularly reviewed and updated to ensure accuracy and completeness.
Test	23	The model has been tested for performance on extreme values and protected attributes	Short description of performance on extreme values and protected attributes	Data scientist	No testing of model performance on extreme values and protected attributes.	The model has been tested for performance on extreme values or protected attributes, but the testing is ad-hoc or incomplete.	The model has been thoroughly tested for performance on extreme values and protected attributes.	The testing for performance on extreme values and protected attributes is regularly reviewed and updated, and the review process is well-defined and executed.	The organisation has an automated system for continuously testing and monitoring the model for performance on extreme values and protected attributes, and the system is regularly reviewed and updated to ensure accuracy and completeness.
Test	24	Patterns of failure have been identified	FMEA, e.g., error curves, overfitting analysis, exploration of incorrect predictions	Data scientist	No identification of patterns of failure.	Some patterns of failure have been identified, but the process is ad-hoc or incomplete.	Patterns of failure have been systematically identified and documented.	The process for identifying and documenting patterns of failure is regularly reviewed and updated, and the review process is well-defined and executed.	The organization continuously reviews and updates its processes to identify and address patterns of failure.
Test	25	Key failure modes have been addressed	Short description of how to resolve or account for key failure modes	Data scientist	No processes or procedures are in place to address key failure modes in AI systems.	Processes and procedures are in place to identify and address some key failure modes, but they are not consistently applied.	Standardized processes and procedures are in place to address key failure modes. These processes and procedures are documented and followed consistently.	Key failure modes are actively monitored, and processes and procedures are updated as necessary to address new or emerging failure modes. The effectiveness of these processes and procedures is measured and reported regularly.	The organization continuously improves its processes and procedures for addressing key failure modes in AI systems based on ongoing monitoring and analysis.
Test	26	The model fulfils the established performance criteria levels	Documentation of model performance	Project manager	Model performance is not documented.	Model performance is documented on ad-hoc basis.	Model performance is regularly documented.	Performance criteria levels are actively monitored, and deviations are addressed promptly. Performance metrics are reported regularly.	The model performance does not dictate the systems criteria levels. - There are procedures in place to prevent this from happening.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Deploy	27	The deployment strategy has been documented	Short description of the deployment strategy	Product owner	No deployment strategy is in place.	A deployment strategy for AI systems has been developed, but it is not consistently applied or	A deployment strategy has been developed, and it is consistently applied and formally documented.	The deployment strategy is regularly reviewed, updated, and tested.	The deployment strategy is continually reviewed, updated, and tested based on feedback and changes in the environment.
Deploy	28	The serving strategy has been documented	Short description of the serving strategy	Product owner	There is no serving strategy documented.	A serving strategy is developed, but it is not consistently applied or formally documented.	A erving strategy has been developed, and it is consistently applied and formally documented.	The serving strategy has been implemented and monitored for effectiveness. - Any deviations from the serving strategy are documented and reviewed.	The serving strategy is periodically reviewed and updated to ensure alignment with the organization's goals and external developments.
Deploy	29	The risks associated with the given serving and deployment strategies have been identified	Short description of identified risks	Product owner	There is no process to identify risks associated with the serving and/or deployment strategies.	The process to identify risks associated with the serving and/or deployment strategies is informally documented. - Risks are	The process to identify risks associated with the serving and/or deployment strategies is formally documented and consistently applied.	The risk identification proces is integrated in the deployment proces and regularly reviewed and updated.	The risk identification proces is continuously monitored and assessed to keep up with changes in the system and environment.
Deploy	30	The risks associated with the given serving and deployment strategies have been addressed	Short description of how to resolve or account for key risks	Product owner	No risks have been addressed.	Risks are adressed on an ad-hoc basis.	There is a formal proces for risk mitigation which is consistently applied.	Risk mitigation plans are periodically reviewed and updated. Effectiveness of mitigation measures is monitored.	Continuous monitoring and improvement of risk mitigation plans and strategies to ensure compliance with changing regulations or ethical standards.
Deploy	31	The model fulfils the established performance criteria levels in the production environment	Performance in the production environment	Product owner	Model performance is not documented.	Model performance is documented on ad-hoc basis.	Model performance is regularly documented.	Performance criteria levels are actively monitored, and deviations are addressed promptly. Performance metrics are reported regularly.	- The model performance does not dictate the systems criteria levels. - There are procedures in place to prevent this from happening.
Sustain	32	The risks associated with changing data quality and potential data drift have been identified	A short description of the risks associated with data quality is captured (e.g., data drift, bias drift, feature attribution drift)	Product owner	There is no process to identify risks associated with changing data quality and potential data drift.	The process to identify risks is informally documented. - Risks are identified on an ad-	The process to identify risks is formally documented and consistently applied.	The risk identification proces is integrated in the operation proces and regularly reviewed and updated	The risk identification proces is continuously monitored and assessed to keep up with changes in the system and environment.
Sustain	33	The risks associated with model decay have been identified	A short description of the risks associated with model decay is captured	Product owner	There is no process to identify risks associated with model decay.	The process to identify risks is informally documented. - Risks are identified on an ad-	The process to identify risks is formally documented and consistently applied.	The risk identification proces is integrated in the operation proces and regularly reviewed and updated.	The risk identification proces is continuously monitored and assessed to keep up with changes in the system and environment.

Sub-Domain		Item	Support	Respondent	1: Initial	2: Repeatable	3: Defined	4: Managed and Measurable	5: Continuous improvement
Sustain	34	The strategy for monitoring and addressing risks associated with data quality and drift; and model decay has been defined	Outline of monitoring strategy (e.g., error classification, critical threshold values for data drift and model decay)	Product owner	No risks have been addressed.	Risks are adressed on an ad-hoc basis.	There is a formal proces for risk mitigation which is consistently applied.	Risk mitigation plans are periodically reviewed and updated. Effectiveness of mitigation measures is monitored.	Continuous monitoring and improvement of risk mitigation plans and strategies to ensure compliance with changing regulations or ethical standards.
Sustain	35	Periodic reviews of the AI applications with regard to the ethical values have been set	Review schedule and format	Top manager	There are no reviews in place.	AI applications are not completely or on an ad-hoc basis reviewed.	There is a formal and periodic proces for reviewing the AI applications with regard to the ethical values.	The proces for reviewing the AI applications is periodically reviewed and updated.	Reviews are integrated with the organization's quality management system and continuously improved.
Maintain	36	The organisation has a strategy for how to update the AI application continuously	Frequency of updates and documentation of model changes	Product owner	There is no update strategy defined.	- A strategy for updating the application is in place, but it is not comprehensive. - Model changes are nor formally documented.	A comprehensive strategy for updating the application is in place and periodically reviewed. - model changes are formally documented.	The strategy is integrated with the organization's change management framework. - The strategy is tested periodically to ensure effectiveness.	The strategy is continuously improved based on feedback and analysis of incidents, and includes a plan for retiring the AI application when necessary.
Maintain	37	A complaints process has been established for users of the AI system to raise concerns or suggest improvements	Short description of the complaints process (e.g., point of contact)	Product owner	No complaints process is in place.	An informal complaints process is established, but it is poorly communicated to users.	A complaints process is formally established and clearly communicated to users.	The complaints process is regularly reviewed and improved based on feedback from users.	The complaints process is integrated into the organization's overall continuous improvement efforts.
Maintain	38	A problem-to-resolution process has been defined	Outline of problem-toresolution process	Product owner	No problem-to-resolution process is defined.	A problem-to-resolution process is defined, but not consistently used or documented.	The problem-to-resolution process is formally documented and followed for significant problems.	The problem-to-resolution process is consistently used, and resolutions are documented.	The problem-to-resolution process is actively monitored and reviewed for improvement.
retirement		The risks of decommissioning the AI system have been assessed	Documentation of decommissioning risks	Product owner	The risks of decommissioning the AI system have not been assessed.	The risks of decommissioning the AI system are assessed on an ad-hoc basis and not formally	The risks of decommissioning the AI system have been assessed and are formally documented.	The risks of decommissioning the AI system are periodically reviewed and updated as needed.	The proces of assessing risks of decommissioning the AI system is periodically reviewed and updated.
retirement		The strategy for addressing risks associated with decommissioning the AI system	Outline of the strategy to manage the risks of decommissioning AI (e.g., data residuals: what will happen to data records, model accessibility and interfaces to other systems)	Top manager	No strategy for managing risks associated with decommissioning the AI system is in place.	A strategy for managing risks associated with decommissioning the AI system is in place, but it is not documented.	A strategy for addressing risks associated with decommissioning the AI system is in place and is formally documented.	The strategy for addressing risks associated with decommissioning the AI system is periodically reviewed and updated as needed.	The strategy for addressing risks of decommissioning the AI system is periodically reviewed and updated.

Appendix C

Review questionnaire

This section displays the contents of the questionnaire as used in the Review interviews. The questionnaires closed questions are questions from the System Usability Scale (SUS). The open questions are a simplified and summarised version of the Technology Acceptance Model (TAM model). Combining these questions should give a quantified view of the perceived usefulness of CompAI by the respondents.

Closed Questions

Instructions: For each of the following statements, mark one box that best describes your reaction to CompAI

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	○	○
2. I found the system unnecessarily complex.	○	○	○	○	○
3. I thought the system was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this system.	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
8. I found the system very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the system.	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
10. I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○
11. I found the home dashboard and graphs to be usefull	○	○	○	○	○
12. I found the self-assessment module to be usefull	○	○	○	○	○
13. I found the Summary Data Sheet for the EU database module to be usefull	○	○	○	○	○
14. I found the External scorecard module to be usefull	○	○	○	○	○

Combined Questions

Instructions: For each of the following statements, mark one box that best describes your reaction to CompAI and explain your choice.

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
Using this product would make it easier to do my job.	○	○	○	○	○

Please explain your choice:

-

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
It would be easy for me to become agile with the product.	○	○	○	○	○

Please explain your choice:

-

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
CompAI's capabilities meet my requirements.	○	○	○	○	○

Please explain your choice:

Appendix D

Exported SDS

Appendix D shows an example of the Summary data sheet (SDS) exported to PDF. The SDS includes all necessary information for registration in the EU AI Database as defined by CapAI. This document could be used for internal documentation and for submission to the EU's AI Database.

CompAI

Summary Data Sheet
To be submitted to the EU's database
System: Test
Date: 2023-06-21

Summary Data Sheet - Test

1. Contact details provider.

Name, address and contact details of the provider.

CompAI, Leiden, compai@email.com, 071-4034395

2. Alternative contact details provider.

Where another person carries out submission of information on behalf of the provider, the name, address and contact details of that person.

Ms Smith, CTO of Enterprise Inc., concern@EnterpriseInc.com

3. Contact details authorised representative.

Name, address and contact details of the authorised representative, where applicable.

Ms Smith, CTO of Enterprise Inc., concern@EnterpriseInc.com

4. System details.

AI system trade name and any ambiguous reference allowing identification and traceability of the AI system.

CompAI, CompliantAI

5. System description.

Description of the intended purpose of the AI system.

Simplifying AI Act Compliance

6. Status.

Status of the AI system (on the market, or in service; not placed on the market/in service, recalled).

not placed on the market

7. Certificate.

Type, number and expiry date of the certificate issued by the notified body and the name of identification number of that notified body (where applicable).

v1.0.0-beta, 071, May 2019, EU AI Board

8. Certificate copy.

A scanned copy of the certificate referred to in point 7 (where applicable).

n.a.

9. Member states list.

Member States in which the AI system is or has been placed on the market, put into service or made available in the Union.

Netherlands

10. Conformity declaration.

A copy of the EU declaration of conformity referred to in Article 48.

See Appendix

11. Instructions of use.

Electronic instructions for use; this information shall not be provided for highrisk AI systems in the areas of law enforcement and migration, asylum and border control management referred to in Annex III, points 1, 6 and 7.

Simplifying AI Act Compliance

12. Additional information.

URL for additional information (optional). Providing this link is optional, yet in our view it is useful to include it here as well as in the external scorecard, which we are proposing below as an additional document to be made available publicly.

See Appendix

Appendix E

Exported ESC

Appendix D shows an example of the External Scorecard (ESC) exported to PDF. The ESC includes the four essential elements defined by CapAI. This document could be used for internal documentation and for external publication to stakeholders.

CompAI

LOGO HERE

External Scorecard

To be published to external stakeholders

System: CompAI

Date: 2023-06-21

External Scorecard - CompAI

1. Purpose
The CommendIXAI system is a recommender system that analyses past purchases and browsing data. It seeks to improve the services and products we recommend when contacting our customers, in order to provide tailored offerings that provide maximum value to our customers

2. Values
Our guiding values at Enterprise Inc are: * Fairness * Transparency * Inclusion A detailed description is available here: www.enterpriseInc.com/values

3. Data
We use proprietary and private data No externally sourced data is used. Consent has been obtained in compliance with GDPR. Protected variables are used (gender and age).

4. Governance
Ms Smith, CTO of Enterprise Inc., is overseeing our AI systems. Complaints and concerns can be raised with her via: concern@EnterpriseInc.com Date of initial deployment: May 2019; last updated: June 2021; Next regular update: June 2022.

Appendix F

CompAI

This section consists of screen captures of all relevant pages of the CompAI system.

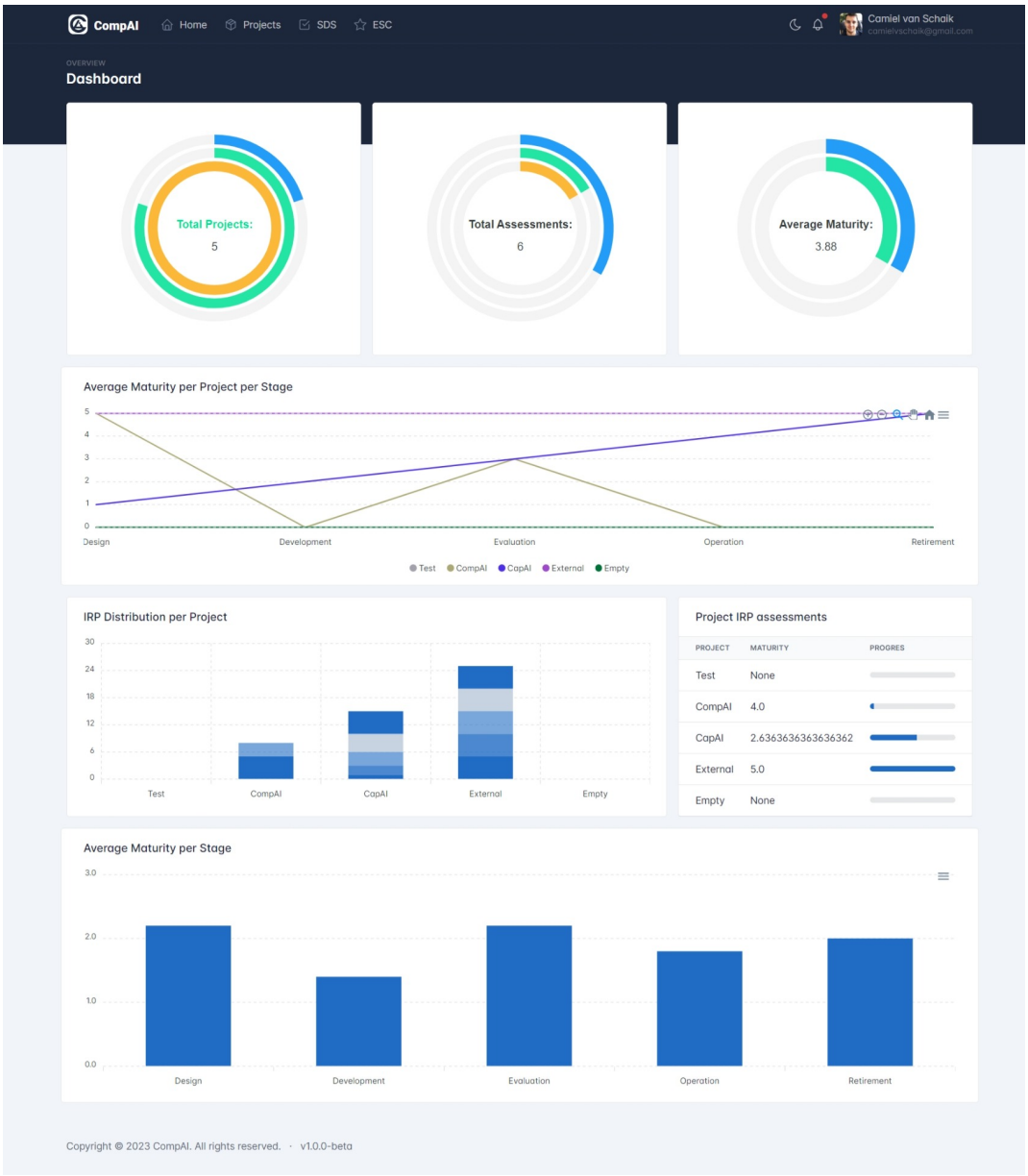


Figure F.1: The CompAI dashboard which features numerous charts and visualisations of data regarding projects registered in the CompAI system. The purpose of the dashboard is to provide the user with useful insights about their status regarding AI Act compliance and tools to communicate these insights to stakeholders.

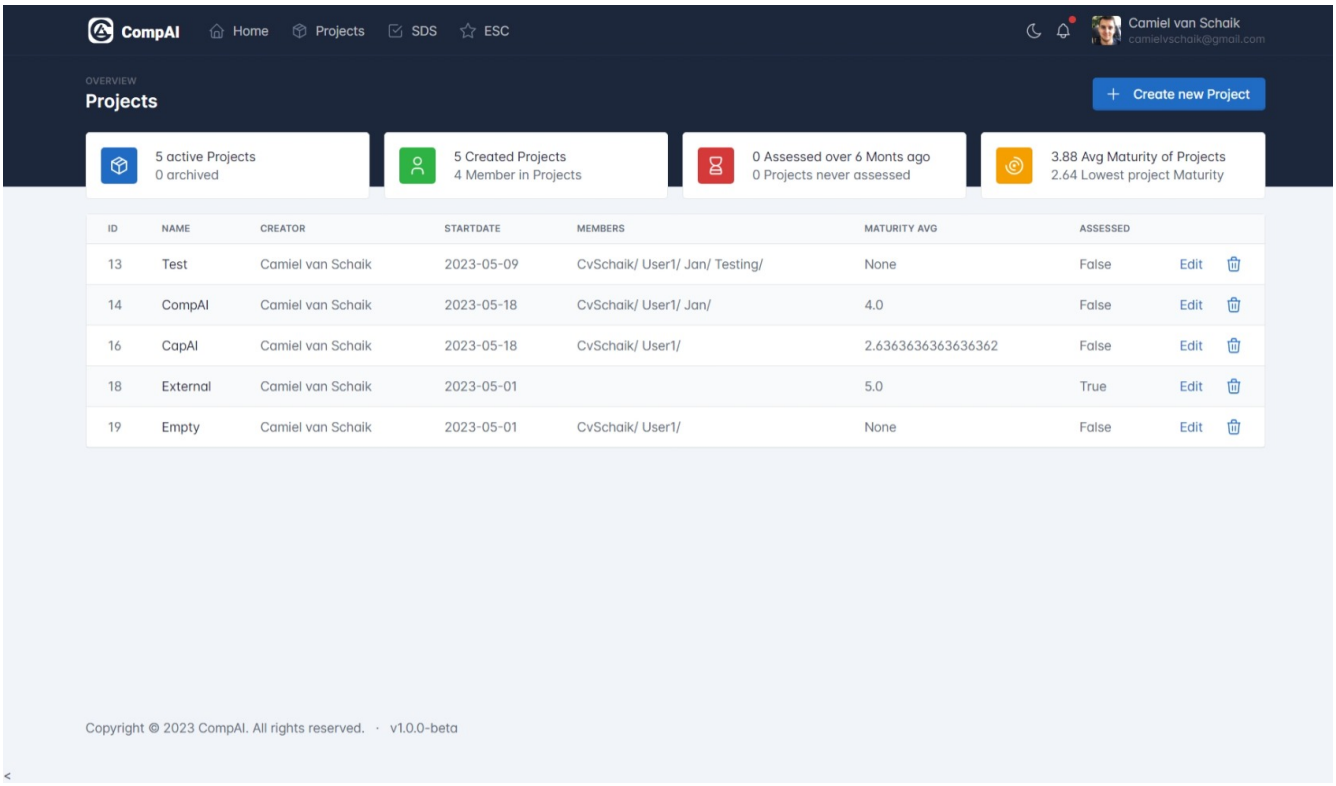


Figure F.2: The CompAI project page consists of a list of all registered projects and their statistics. The page provides users with a quick overview of the status of all projects to ease the prioritization of projects.

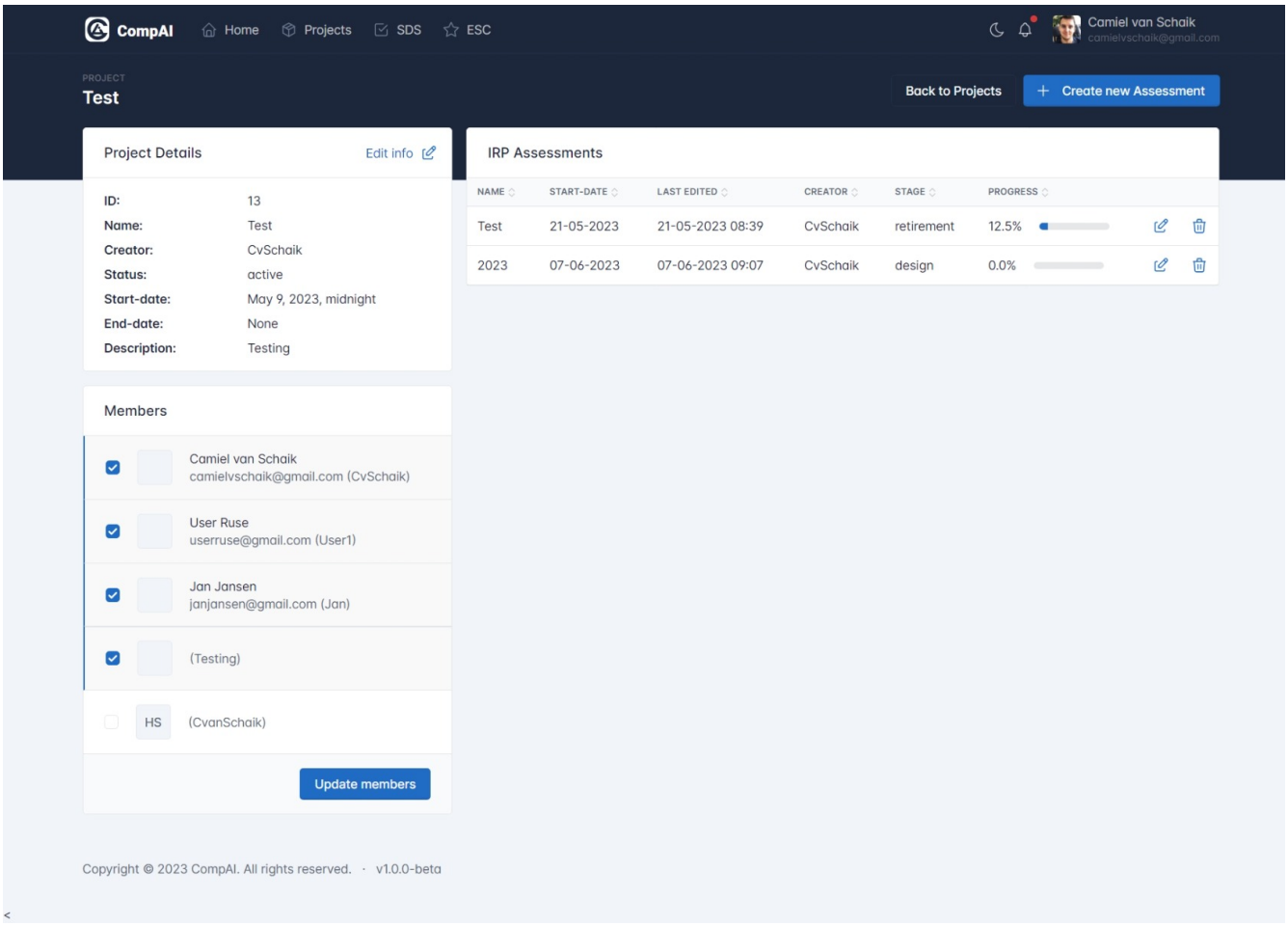


Figure F.3: The CompAI project detail page which shows the user relevant details of a certain project. The page gives the possibility to modify the project's data and create, edit and delete IRP Assessments for a specific project.

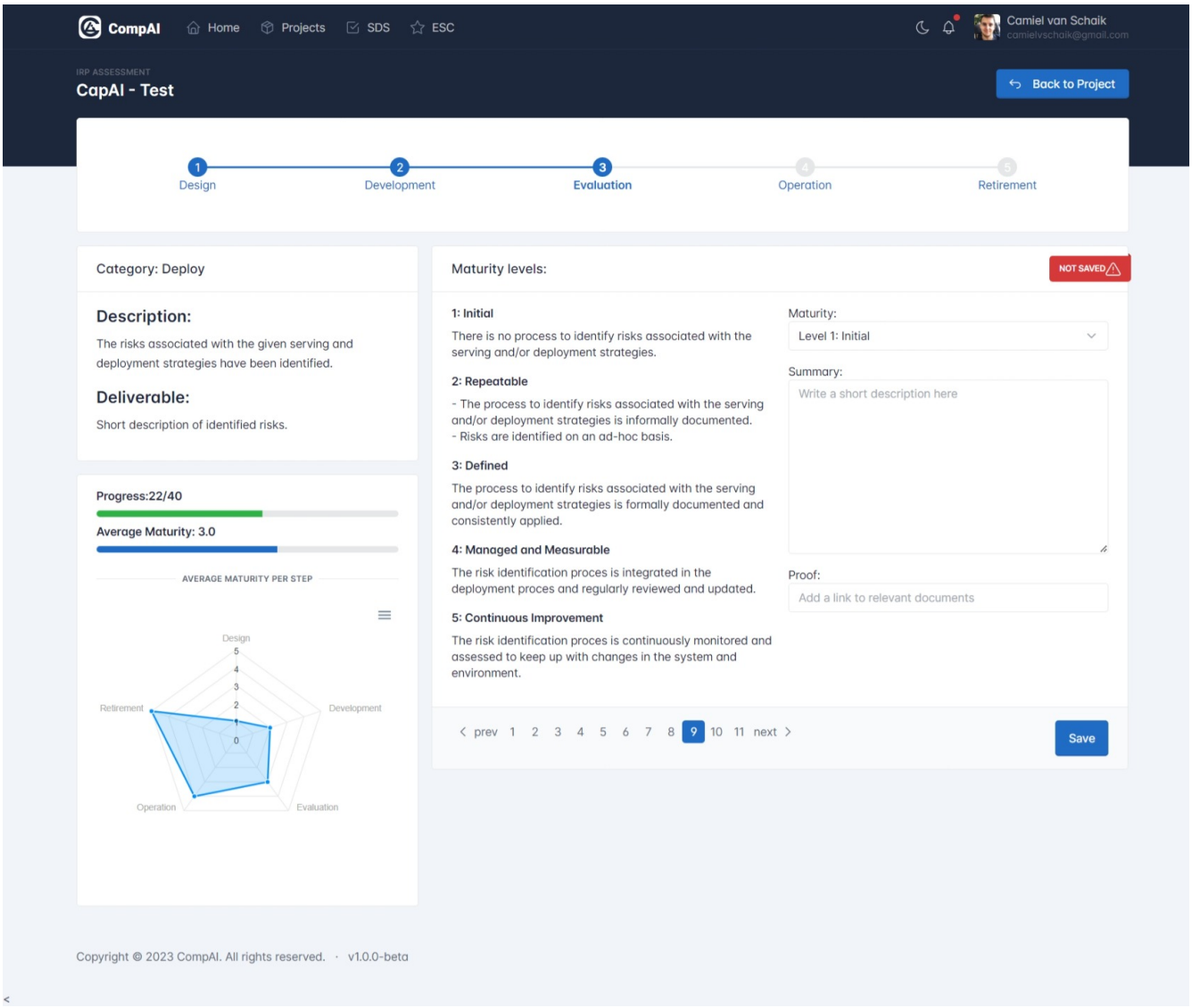


Figure F.4: The CompAI IRP assessment page. This page combines the CapAI IRP with our maturity model. Users can fill out the maturity form for every IRP item and get instant feedback with real-time statistics. Using the timeline the user can navigate through different stages of the AI lifecycle. The red banner denotes if the current item is saved and turns green when it is.

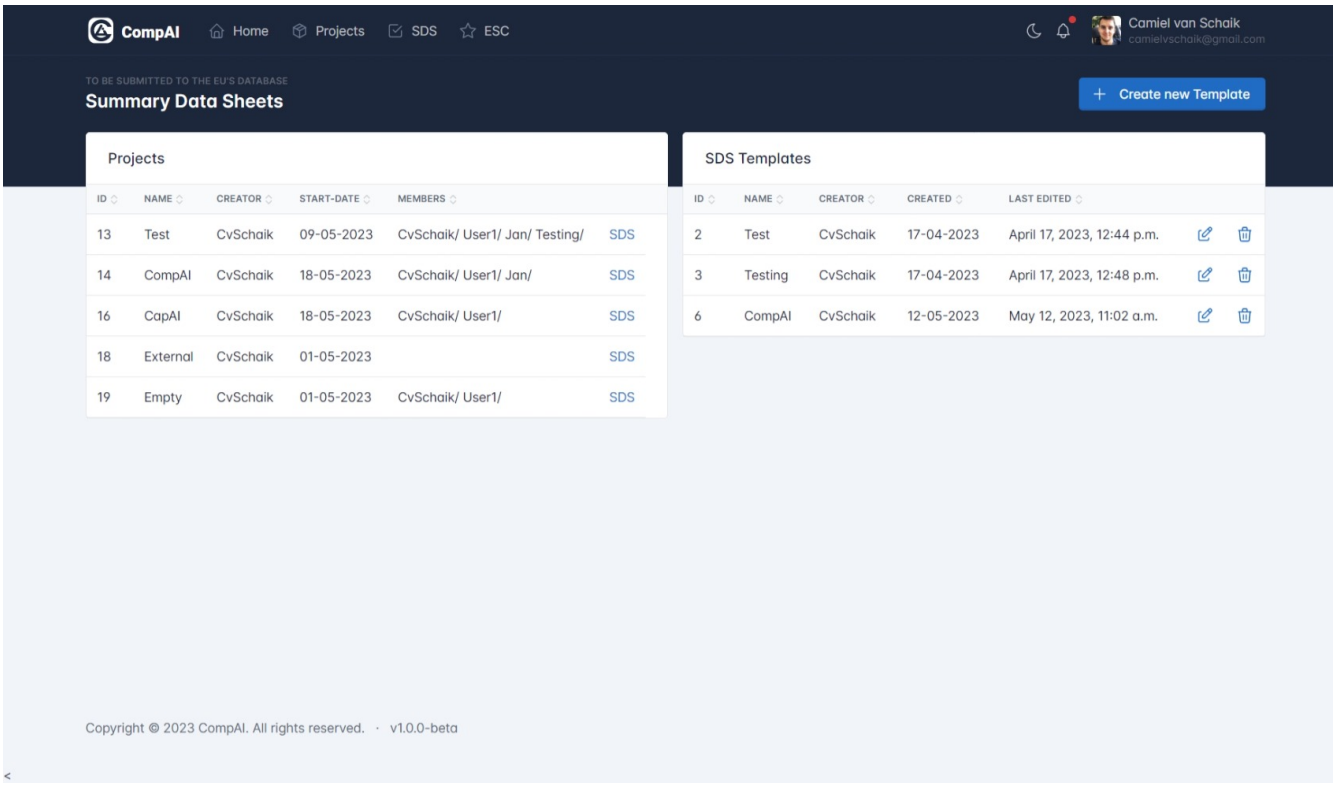


Figure F.5: The CompAI SDS overview page. This page offers an overview of all projects and SDS templates. Selecting either a project or template redirects the user to the page for filling out the SDS form for that project or template respectively.

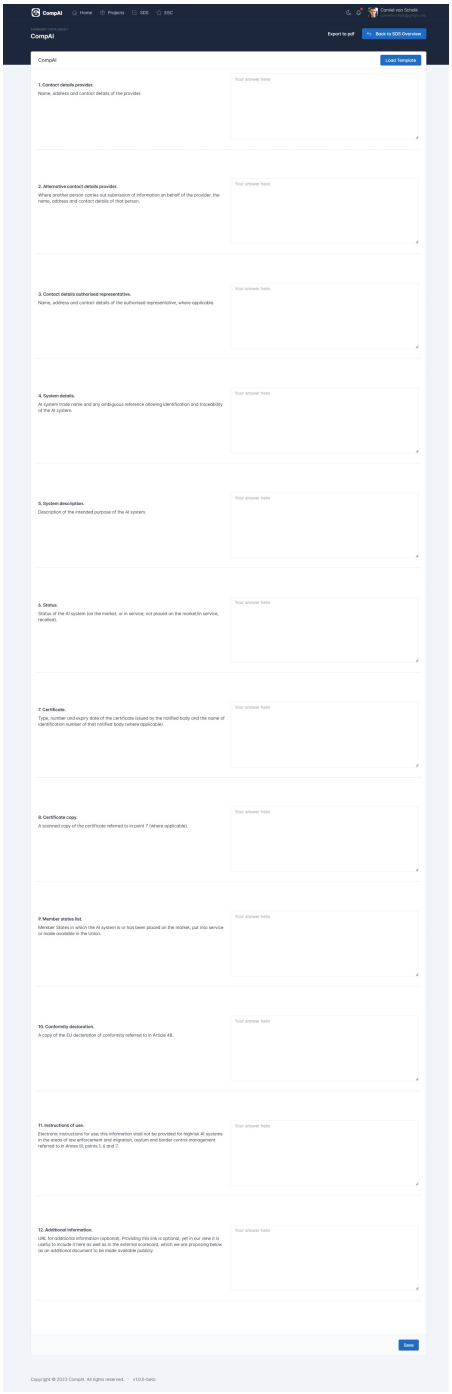


Figure F.6: The CompAI SDS page This page Functions as a form for filling out a project's SDS. Using the buttons at the top the user can either import an SDS template or export the SDS as a PDF.

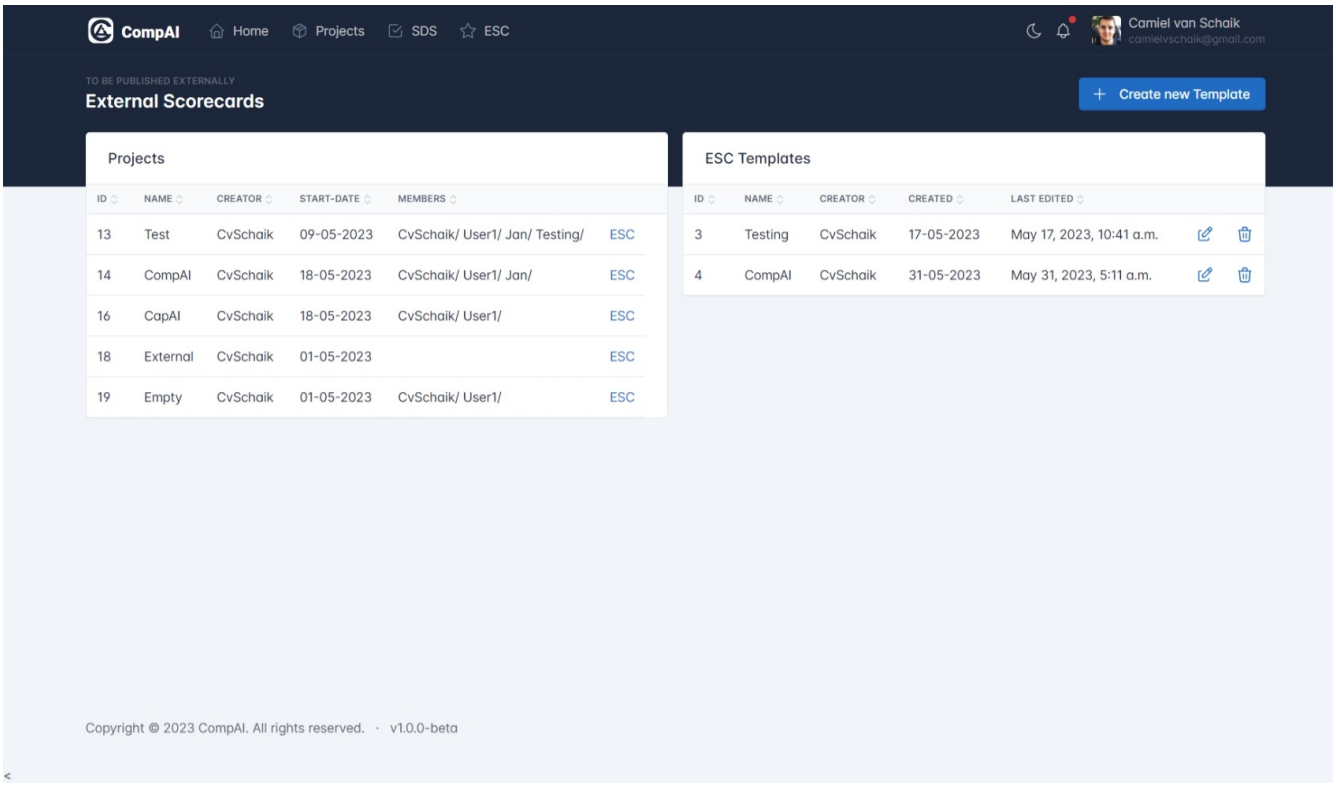


Figure F.7: The CompAI ESC overview page. This page offers an overview of all projects and ESC templates. Selecting either a project or template redirects the user to the page for filling out the ESC form for that project or template respectively.

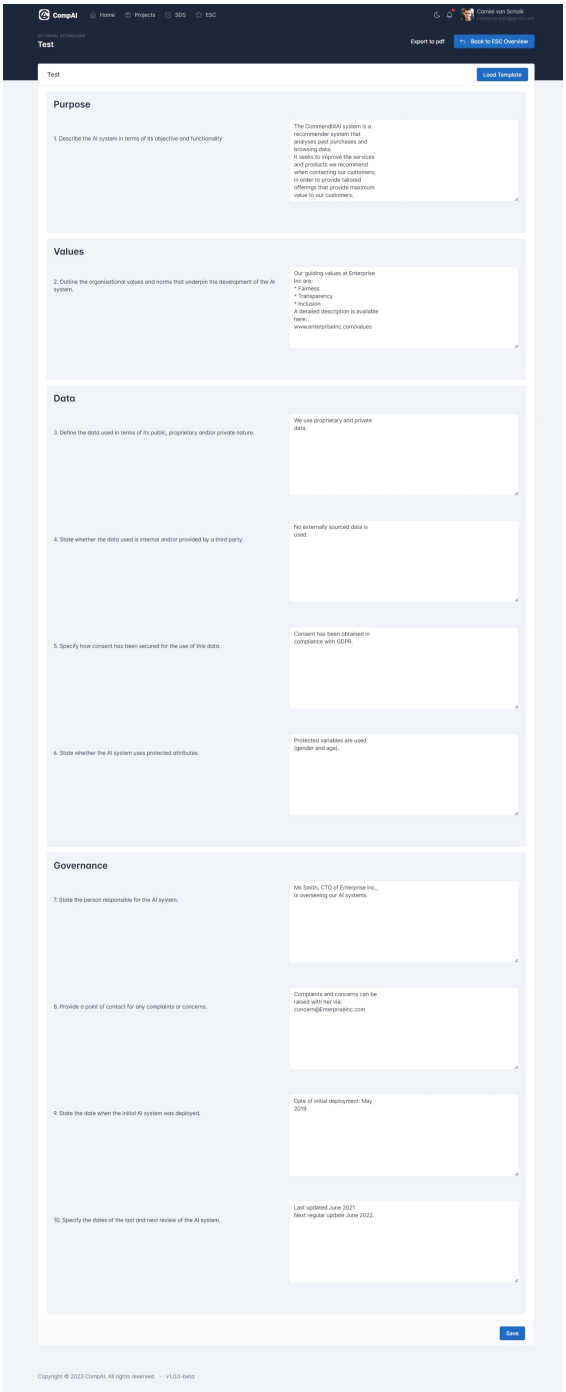


Figure F.8: The CompAI ESC page. This page Functions as a form for filling out a project's ESC. Using the buttons at the top the user can either import an ESC template or export the SDS as a PDF. The form consists of all relevant elements from the CapAI ESC. All form elements are grouped per ESC element to give the user a better overview of the form.