Master of Science in Business Information Technology

The Regulatory Situation of Artificial Intelligence in Switzerland

Ethical and Fairness Considerations and the Implications for Innovation

---

MASTER THESIS

Iris Amrein

Supervisor
Dr. Donnacha Daly

Expert
Prof. Dr. Orlando Budelacci

University
Hochschule Luzern

Lucerne
May 21, 2022

Master Thesis at Lucerne School of Computer Science and Information Technology

Title: The Regulatory Situation of Artificial Intelligence in Switzerland

Student: Iris Amrein

Degree Program: MSc in Business Information Technology

Final Year: 2022

First Reviewer: Dr. Donnacha Daly

Second Reviewer: Prof. Dr. Orlando Budelacci

Coding / Classification of the work: public

Affidavit attesting sole authorship

I hereby declare that this thesis is entirely my own work, produced without any unauthorized third-party assistance, that I have correctly cited all sources, literature and other aids used, that I have correctly and appropriately identified passages taken verbatim or in terms of content and that I have taken into account the client's confidentiality and respected the copyright regulations of the Lucerne University of Applied Sciences and Arts.

Lucerne, May 21, 2022

Submission of this thesis onto the university portfolio database: Student Confirmation

I confirm that I have correctly uploaded and saved the master thesis onto the university portfolio database in accordance with the instructions sheet. I have handed in all relevant authorizations and acknowledge that no further changes or uploads can be made.

Lucerne, May 21, 2022

Thanks:
I would like to express my gratitude to my primary supervisor, Dr. Donnacha Daly, who guided me throughout this project. His insightful feedback pushed me to reach the best possible version of this thesis.

I also wish to extend my special thanks to all survey participants and interview partners who helped me gain deeper insights on the topic.

In addition, I thank my friends and family who have supported me throughout the process of writing this thesis.

Abstract

Although Artificial Intelligence applications provide many benefits to companies in many different industries, there is a wide set of examples of AI applications harming individuals, for example through discriminatory decisions. To prevent this harm, many countries have started drafting AI regulations, with the EU AI Act being the first Act to provide a unified regulatory base for a large geographical region. Despite these developments, Switzerland remains in a reserved position regarding AI regulation. Given that innovation is important for Switzerland and there are many AI companies and startups, this thesis aims to recommend actionable practices for governmental and industry decision makers to build an effective and innovation-fostering AI regulation. While Switzerland's technology-neutral approach to regulation certainly has advantages to keep up with rapidly evolving technologies, companies are still operating with a large amount of legal uncertainty. Most surveyed industry players favor a more specific approach to AI regulation. The thesis proposes two frameworks, one for governmental regulation and one for self-regulation, that help mitigate this legal uncertainty while still allowing for innovation to thrive. The governmental regulatory framework contains actions to regulate AI applications, inform industry players about regulation and monitor the compliance with the regulation. The self-regulatory framework on the other hand builds upon the governmental basis and derives ethical principles as well as actionable practices for each phase in the software development process. Together, the frameworks contribute to overcoming the most threatening challenges of AI and therefore minimize AI harm, all without hindering innovation of AI technology.

Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence

API Application Programming Interface

BWC Biological Weapons Convention

CKD-EPI Chronic Kidney Disease Epidemiology Collaboration

CNAI Competence Network for Artificial Intelligence

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

CSO Civil Society Organization

CSR Corporate Social Responsibility

DLT Distributed Ledger Technology

EU European Union

FaST Fast Screening Tool

GDPR General Data Protection Regulation

GPT3 Generative Pre-trained Transformer 3

HIV Human Immunodeficiency Virus

IoT Internet of Things

kNN K-Nearest Neighbor

NIST National Institute of Standards and Technology

NPO Non-Profit Organization

OECD Organisation for Economic Co-operation and Development

PIPL Personal Information Privacy Law

ROS Risk-oriented Sanctions

SATW Swiss Academy of Technological Sciences

UN United Nations

US/USA United States of America

1 Introduction

Over the past half-century, AI has developed into a general technology that finds application in almost all industries today, from farming to banking. But as new beneficial use cases arise, so do news about scandals and malicious uses of AI. For example, health care algorithms are exhibiting racial bias and deep fakes are used in (political) modern warfare. Under self-regulatory measures, some well-known and successful companies have halted development on their AI products or taken action to prevent them from doing harm. Amazon, Microsoft and IBM for example have stopped selling their facial-recognition algorithms to law enforcement[1][2], and OpenAI has changed its distribution model for GPT3 from open-source to commercial API, in order to protect the algorithm from being abused [3].

While some companies seem to recognize the risk their applications pose and offer solutions to minimize it, other companies are less active. Clearview.ai for example still owns a database full of pictures of faces, which it sells to law enforcement [4].

So the question is how can AI be prevented from doing harm without missing out on the benefits the technology provides? Some governments are already forming solutions to this question by releasing AI regulation drafts or rules. In April 2021, the EU has released a draft for an International AI Act that aims to provide a unified set of rules for the participating nations. But where does Switzerland stand in this regard and what can it do to answer the question at hand?

This thesis aims to answer this question by evaluating the strengths and weaknesses of Swiss AI regulation from the industry's perspective and in comparison with other countries. It will also identify opportunities and threats for Swiss AI regulation to propose further actions. It does so by conducting literary research of Switzerland and other countries' AI regulations and by surveying industry decision makers and AI regulation experts.

The thesis will start with the identification of the research gap, followed by the description of the research questions and the research design to answer them. Then it will take a deeper look into the topic of Artificial Intelligence and its challenges before building an understanding of the topic of regulation. After building this theoretic basis, the thesis analyzes the current situation of AI regulation in Switzerland and leading AI countries as well as which learnings can be derived from the regulation of other disruptive technologies. Subsequently, the thesis paints a picture of the perspective from industry players and AI regulation experts. All insights from the literature review and primary research are then combined to answer the research questions. Finally, the thesis ends with the description of a framework for AI regulation that aims to mitigate AI harm while still encouraging beneficial development.

2 Literature Overview

In this chapter, the current state of national and international AI regulation studies will be elaborated. First, the research process is briefly explained before national and international AI regulation studies are introduced. Finally, the research gap is concluded.

Research Procedure. In order to find scientifically accurate papers and studies that discuss AI regulation on a national and international level, the following keywords were defined, whereas the term "Artificial Intelligence" was also replaced with its abbreviation "AI" to find all possible results:

- Artificial Intelligence regulation study
- Artificial Intelligence national regulation
- Artificial Intelligence Policy review
- Local Artificial Intelligence regulation report
- National Artificial Intelligence regulation report
- International Artificial Intelligence regulation report
- Artificial Intelligence regulation guidelines

As a starting point for the search, the keywords listed above were entered into the Google Scholar search engine. For results whose title, publisher, publication date and description seemed to correspond to the topic at hand, the article was evaluated for its relevance, citation eligibility and citation worthiness. This resulted in a list of 15 potentially relevant papers and articles.

The Role of Policymaking in AI Success. The research community agrees that policymakers[1] need to divert their attention to the challenges raised by AI and to take the opportunity to decide upon how this technology will be shaped [5][6][7]. Madhavan et al., Thierer et al. and Lauterbach say that policy decisions will influence to what extent the benefits of new technologies can be reaped and will therefore lead to long-term consequences for business and individuals alike [8][9][10]. Aside from consequences regarding what can be done with AI, Smuha and Madhavan et al. explain that regulations can also strengthen the public's trust in AI [6][8].

[1]The literature referenced does not define what they see as "policymakers". For the purpose of this study, the term "policymakers" encompasses every decision-maker on a government level.

Reactive vs. Proactive Policymaking. While the research community agrees that policymaking will have an impact on AI's success, it is divided about what an ideal policymaking process should look like. Thierer et al. argue that a proactive approach would harm the field of AI because it would hinder beneficial development, as few innovators will want to work on projects that are at risk to violate a strict regulatory regime. They further state that many companies and engineers do not want to develop a bad reputation and will therefore try to avoid racist and/or sexist bias as well as corruption in their algorithms on their own without regulatory frameworks. They also mention that the market would regulate itself as when a company draws incorrect conclusions or misses opportunities, a competitor will fill the gap with better analysis [9]. Most authors in the literature overview sample however describe a proactive approach as ideal. The advantages of a proactive approach are described as strengthening the public's trust in AI technology and nudging researchers and developers into a "beneficial" direction [7][8][11][12]. Guihot et al. describe such a proactive approach as not just to pass laws but to influence participants in AI development through signals and expectations, then interact with the relevant industries to be in a well enough educated position to regulate the areas that pose the most risk [13].

National Regulatory Situations. Roberts et al. studied China's "Next Generation AI Development Plan" for regulatory requirements and strategies. They found that China has set a big goal of being the world leader in AI by 2030 and their plan offers a clear intention to define ethical norms and standards. Yet, this plan is too high-level and lacking implementation as well as offering too many loopholes and exceptions for the government to bypass privacy protection. Finally, they stress the importance of China as a central actor in global discussions about AI governance [14]. Madhavan et al. and Gutierrez focused on the United States of America. They acknowledge that AI opens gaps in US policies that need to be addressed. However, they note that these gaps are not specifically created by AI methods but by certain AI applications [8][15]. Binder et al. have analyzed the regulatory situation in terms of AI in Switzerland and compared it to the EU's regulation proposal. They find that there is a need for adaption of laws in Switzerland in order to handle challenges posed by AI but recommend that such rules should not be manifested in a specific AI law but rather an adaption of existing laws that correspond to the respective AI applications [16].

Need for International and Transnational Regulation. Given that Artificial Intelligence is a global issue, with growing worldwide activism and global impact on society, two reports explore proposals for an international regulatory framework. Erdelyi and Goldsmith argue that national law cannot govern such a transnational issue, and Geetha Sethu specifically recommends the United Nations as the right body to regulate AI. She mentions that AI weapons can be used in times of peace but also of war and that AI can impact human rights and humanitarian law hence the UN, which is already governing

these issues, would be the perfect governing body [7][17].

The Role of Self-Regulation. The research community accepts that although there are few regulations, voluntary guidelines and ethical principles do exist in private businesses that work on or with AI. However, they criticize that such voluntary guidelines are not enforceable when they are breached and therefore operationalized regulation is needed[6][18]. The community also agrees that ethics should not be the only aspect regulated in AI, but need to be incorporated into a regulatory framework, as AI may not only affect individual rights but also society as a whole [7][19].

Research Gap. The literature overview shows that little scientific research on AI regulation is conducted with focus on one country's situation. If a paper treats such a topic, it mostly relies on literature analysis of particular documents or laws. The majority of the research covered focuses on AI challenges, then proposes a way to approach regulation of these challenges based on scientific research and proven concepts. There has only been one paper identified which focuses on the situation of Switzerland. As with all of the papers analyzing national situations, this report on Switzerland also does not conduct more than a literature analysis. Most papers identify that regulation will impact research and development in some way, but they do not provide detailed examples on how innovation will be impacted by regulation. Therefore the research gap lies in a scientifically conducted study about what an innovation-friendly Swiss regulatory situation for AI should look like based on inputs from parties that are directly affected by regulations.

3 Research Questions & Research Design

This section describes the Research Questions derived from the research gap and the Research Design to answer those questions.

3.1 Research Questions

In order to address the research gap, the thesis will aim to answer the following research questions:

Q1: What is the current state of AI regulation in Switzerland?

- How does it compare to other regions?
- How does it compare to the regulation of other technologies?

Q2: What is the industry's perspective on AI regulation?

- What is the role of ethics and self-regulation?
- Could technology improvements be an alternative to regulation?

Q3: What are the most important factors to include in a regulation proposition for AI?

- What can be done to protect the public from AI harm?
- How can restriction of beneficial AI development be avoided?
- What impact would an AI regulation have on research & development?

Q4: What are the strengths, weaknesses, opportunities and threats for Swiss AI regulation?

3.2 Research Design

The Research Design represents the process chosen to adequately answer the research questions. The categorization of the research design is followed by a graphical representation and description of the process as well as an explanation of the research methods.

3.2.1 Categorization of Research Design

The answers to the research questions need to be based on data. As the research questions refer to a rather recent topic in a rather small geographical area, existing data is scarce. This prevents the author from doing a secondary analysis of such existing data. Therefore, the goal of the thesis will be to collect and analyze new primary data. As the research questions require a detailed as well as a broad perspective, the data gathering will be conducted in two different forms. Data from companies and institutions affected by AI regulation will be collected through a survey, and data from experts in AI regulation will be gathered by in-person interviews. Therefore, the research procedure follows an exploratory-mixed-methods approach.

Table 1 provides an overview of the research design by categorizing it based on the selected variants and displaying the reason for selecting this variant.

Characteristic of the Research Design	**Selected Variant**	**Reason**
Scientific theoretical approach	Mixed methods approach	Quantitative approach: General understanding about the industry's perspective on the topic; Qualitative approach: Dive deeper into the given answers and evaluate the topic further
Object of study	Empirical study	Independent data collection and analysis
Data basis	Primary analysis	Analysis of the newly collected data
Interest in knowledge	Explorative study	Explore topic with openness for unexpected findings, no hypothesis to validate
Screening site	Field study	Data collection happens in the real world with real world economic players
Number of objects of study	Group study	Analyze several cases across the board, abstract peculiarities of individual cases

Table 1: Categorization of research design including selected variant and associated reason

3.2.2 Graphical Illustration of the Process

Figure 1 displays the most important steps of the research to conduct. The process can be divided into three main parts: A theoretical part with a literature study, a practical part where data is gathered and an analysis part where the information is evaluated and conclusions are drawn. The steps are not fully dependent on each other and a subsequent step can be started before the previous step has been finished. The writing of the thesis is ongoing during the research process and done step by step as soon as new information is found or received.

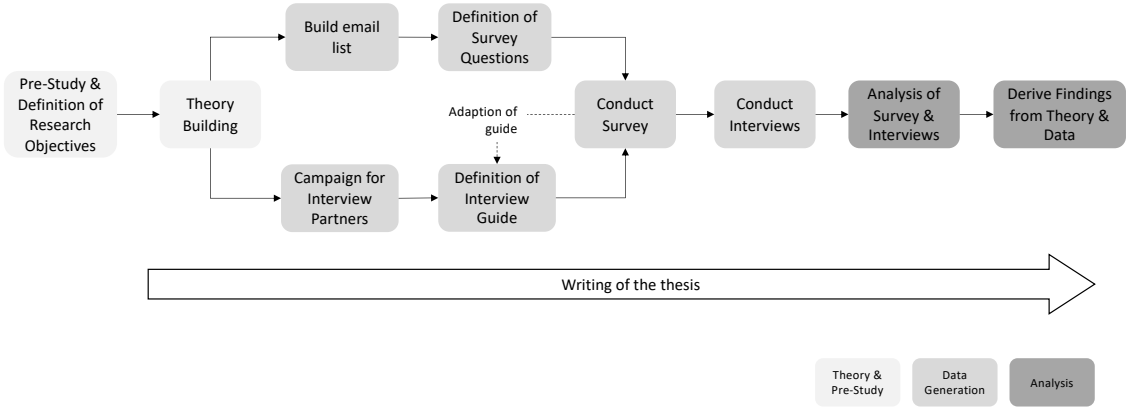


Figure 1: Graphical illustration of the complete research process

3.2.3 Methodology

The research questions show a need for new data, which is gathered by conducting a survey and expert interviews. The data should indicate how the industry views AI regulation and what should be done in terms of AI regulation from the perspective of industry players and AI experts.
Within a limited time frame, a survey can gather more data compared to interviews, as it works with a relatively big sample of sources. Therefore, it is an ideal method to gain an understanding about the industry's general perspective on AI regulation. Interviews however can provide more detailed information and allow for questions to gain a deeper understanding about the topic at hand. The unstructured nature of the method also allows for new topics and perspectives to arise that wouldn't have been caught in a survey. Against this background, a combination of both methods was chosen to gather a diverse overview but also a deeper level of information about AI regulation in Switzerland. The following sections will further elaborate on the methods' execution.

3.2.4 Survey

Sampling. Survey participants should have a good understanding of ethics and regulation in AI and they should be familiar with making decisions about AI applications in their company. Ideal partners would therefore be people that are decision-makers about the application of AI or that are regularly in contact with ethical and/or regulatory questions about AI usage. As a result, the sampling process focused on finding decision-makers on C-Level or members of the extended executive board whose position includes working with AI. An email list of 200 qualified people was compiled through a search for contacts on LinkedIn.

Survey type. In order to make it as convenient as possible to fill out the survey, it was

available online and consisted of mostly closed questions. This typifies it as a mostly standardized online survey.

Survey Set Up. The contacts from the email list were invited to the survey by email or LinkedIn message. Along with the survey link, the possible participants also received a one-pager that described the thesis, its goal and the author in order to provide context (see Appendix A). The survey was set up to only take 10 minutes at the most to complete and was displayed in a one page layout in order to make its completion quick. The survey questionnaire can be found in Appendix B. At the end of the survey, participants were able to choose to enter their email addresses in order to receive the final thesis. The survey was accessible from February 22 to March 31, 2022.

Analysis. The analysis of the survey data was done in Excel. The data was analyzed by applying descriptive statistics. The author refrained from using inferential statistics, as the descriptive statistics method is sufficient to describe the perspective of the participants and identify differences and commonalities.
The analysis was split into three steps:

- Data cleaning
- Sample description
- Data inspection & descriptive analysis

In the first step, incomplete data sets and implausible answers were either completed, corrected or deleted from the sample. For the sample description the frequency distribution and mean values were provided on details such as industry or position. Finally, the clean data was analyzed and graphs and tables were provided to display the results of the survey and indicate the commonalities and differences in the answers given.

3.2.5 Expert Interviews

Sampling. The interview partners should be familiar with aspects of regulation, including AI regulation. Expertise in the impact of regulation on technology in general would be beneficial. They should be able to provide detailed information about the impact and possibilities of AI regulation. Such experts were found in the network of the author and through contacting authors of scientific articles regarding AI regulation in Switzerland. In the end, six interviews were conducted.

Interview Type. The interviews were semi-structured to improve flexibility and scope of the data. The interview guide (see Appendix D) proposed questions and their order but individual adaptions like skipping or deepening certain questions were allowed. The interviews took between 30 and 45 minutes.

Interview Set Up. The participants were invited to the interview by email. In order to provide a quick overview of the thesis for possible participants, the invite contained a one-pager that introduced the topic, the goal and the author (see appendix A). All interviews were held via video call and all interviewees consented to being recorded for the purpose of completing the interview notes afterwards. They were informed that the video and audio files would be fully deleted once this was achieved.

Transcription. In accordance with this thesis' supervisor and due to time constraints, the notes of the interview served as the basis for the analysis instead of a full word-for-word transcript. To exclude errors and biases, the notes were compared and if applicable complemented with the content of the recording of the interview. The notes are kept confidential and are made accessible to the supervisors of the thesis in a separate document.

Analysis. The interviews were analyzed using the qualitative method of category building. The overall goal was to identify commonalities and differences in the interviewees' statements. The analysis included the following steps, which are the central elements of all qualitative data analysis [20, pp. 603f.]:

- Case related evaluation
- Segmentation and Analysis
- Coding
- Categorization

In the **case related evaluation** one interview represents one case. The interview notes were reviewed one after another to develop an initial rough understanding of the obtained data. The **segmentation and analysis** divided the interviews into logical text units. The interview guide hereby served as a starting point to find similar logical text units in the different interviews. Each unit of analysis was then examined in terms of its content. After all units were analyzed, they were assigned a **code** which described the content in a short phrase. Finally, similar codes were gathered into **categories**, which represent thematic content areas.

4 Artificial Intelligence

This chapter introduces the topic of Artificial Intelligence. It provides a definition to distinguish it from human intelligence and other more basic types of IT applications. Subsequently, it displays the different subtypes and components that play a core part in making machines "intelligent", such as Machine Learning and Deep Learning.

4.1 What is AI? - Definition & History

There is no widely accepted definition of the term “Artificial Intelligence”. This section will therefore present a definition for “Artificial Intelligence” for the context of this thesis. After defining the term, the section will then briefly dive into AI’s history to display an overview of the technology’s developments and challenges.

4.1.1 Intelligence - Delimitation of human & Artificial Intelligence

There is currently no way of defining intelligence without putting it into the context of human intelligence [21]. Therefore, this section will seek to understand what distinguishes artificial from human intelligence as a prerequisite to define the term “Artificial Intelligence”.

Definitions of human intelligence often describe it as the (mental) ability to acquire knowledge and apply it to one’s environment in order to resolve a specific problem. The acquisition of knowledge happens through learning from experiences and examples. The application of knowledge often does not stay in one domain, as humans are able to apply what they have learned through solving a problem to a new context without much effort [22][23].

Where human intelligence is very rich in form and function for example through creativity and humor, Artificial Intelligence seems to have only one specific aim: Its goal is to solve problems and thus provide value in the absence of human supervision. This is achieved through learning how to execute complex tasks such as driving cars, performing surgery, or recognizing tumors in pictures [24][25]. It is essential to understand that the goal of AI is not to simulate human intelligence but to focus on solving problems. Most of the time, this is not done by recreating the way humans would solve the problem, but by analyzing the problem at hand and finding the most efficient way to solve it [21].

Nevertheless, modern implementations of AI do share some characteristics of human intelligence, such as learning from past experiences in order to improve performance [25]. Comparing the two must be done cautiously as the complexity of human intelligence is often underestimated which results in overestimation of what AI can do. The key factor

hereby lies in the ability to apply the acquired knowledge to a new context, which is much easier for humans than it is for machines [26]. Research has therefore already identified two classifications for AI applications - narrow and general AI - to take this issue into account:

Narrow AI can only work in a limited and well-defined task domain. It is not flexible and can therefore not apply the learned knowledge from the original to a new context [27]. On the contrary, *general AI* is able to do so. It can apply the acquired knowledge from what it was trained for to new contexts that are unrelated to the original training context [27]. Most AI applications used today are narrow AI applications. An example for an AI application that is sometimes (inaccurately) being labeled as a general AI application is OpenAI's GPT3 algorithm, which can already solve some language-based tasks without being specifically trained for them [28].

Having established that Artificial Intelligence is not a mere "copy" of human intelligence but aims to solve problems independently in place of humans, this thesis defines the term as follows:

Artificial Intelligence is the ability of a non-biological system to solve problems independently and in the same or better quality and speed as a human. Intelligent machines will also continuously improve their performance of solving said problem by taking into account previous feedback and new insights.

It's important to note that the ability to learn and improve distinguishes AI from conventional software. In conventional development, developers have to write code to tell the program what to do step by step in order to achieve the goal. Artificial Intelligence algorithms will learn how to perform a task and improve on that continuously independently and without developers' interference [29].

4.1.2 AI in History - Summers & Winters

The history of Artificial Intelligence begins with its first mention at the Dartmouth Conference in the summer of 1956. John McCarthy had coined the term, defining it as the "science and engineering of making intelligent machines" [30][31][21]. Two years later, in 1958, he invented the programming language LISP, which allowed programs to modify themselves [32].

In the 1970's, the programming language PROLOG replaced LISP as it provided a simpler and faster way of programming for more complex logic structures. This helped AI to become more efficient. Although use cases for this symbolic-based language were minimal, computer scientists, particularly Marvin Minsky, estimated that eight years after its invention, it would allow creating machines with human-like intelligence. As we know

now, this was a great over-estimation of AI's abilities and led to a massive disappointment and loss of trust in the technology when AI did not deliver on the expectations. This led to the first AI winter, which is characterized as an era in which funding and interest in AI technologies were scarce [32][33].

Almost thirty years after the first weight-learning neural network model, discovered by Frank Rosenblatt in 1950, researchers took up work on neural networks again thanks to the re-discovery of the backpropagation and gradient descent technique that had been around since the 60s [34]. Unfortunately, their attempts failed due to a lack of data to train and computing power to run the algorithms and because it was not possible to train large networks or to convert neural networks into formulas and logic rules[32]. After 25 years of more research in neural networks, the work of scientists finally started to pay off. These very powerful networks were a catalyst for more complex use cases, such as image classification [35].

This short excerpt from the history of AI displays that the technology did not have a straight uphill run to success. As predicted by Gartner Hype Cycles, every new technology passes specific cycles, one of which is the depression cycle. As indicated above, AI went through its first depression cycle when the PROLOG language did not deliver on what was expected of it. This is often referred to as the first AI winter. A second AI winter followed when better-performing and easy to handle desktop computers became affordable. After a successful period with expert systems, it became evident that they were much more demanding to update and more expensive to maintain than the new desktop computers. This led to the downfall of expert systems and the second AI winter [36].

Seeing as cycles are destined to repeat themselves, the question arises whether a third AI winter lies on the horizon. The author deems this possibility relatively low as the success of AI technology is no longer measured as a whole but fragmented into different subtypes and applications. Figure 2 shows that the Gartner Hype Cycle 2021 differentiates several specific applications of AI. Although new winters will likely appear as these applications make their way through the cycle, these winters will not impact the technology but specific disciplines, such as generative AI.

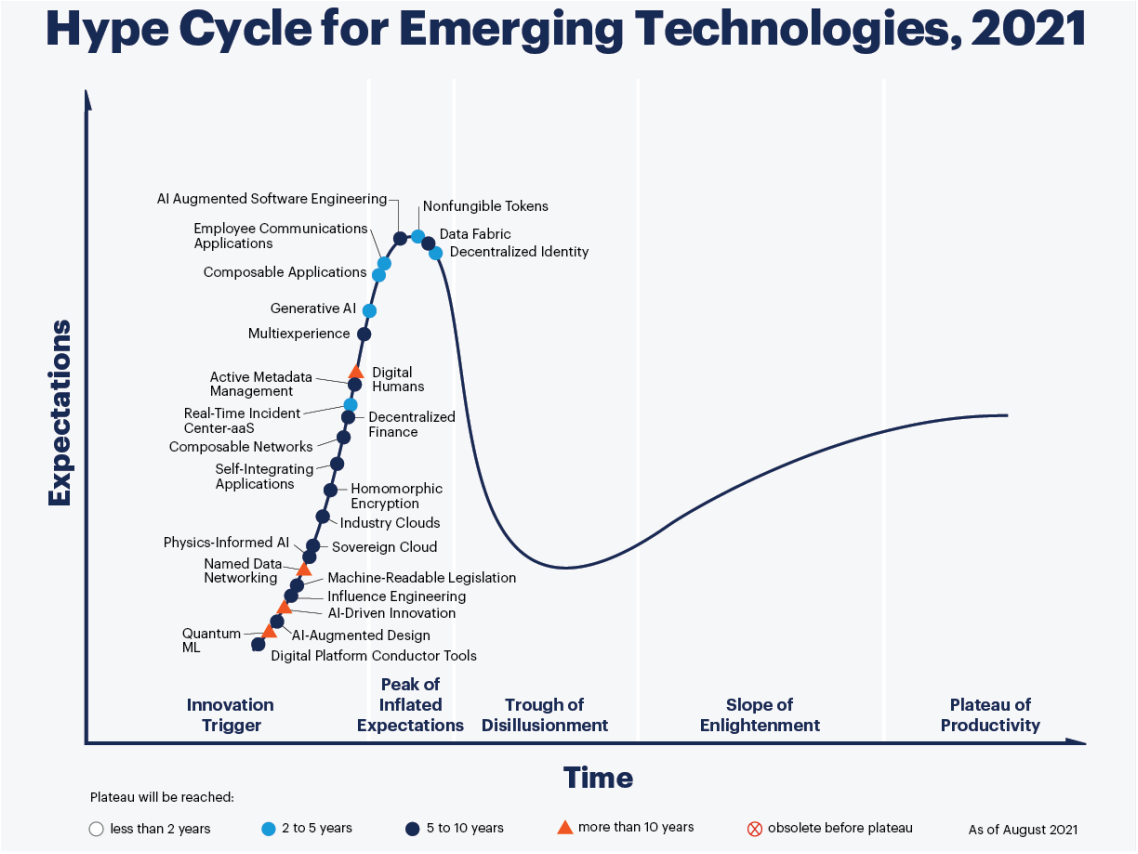


Figure 2: Gartner Hype Cycle for 2021, showing different disciplines of AI [37]

4.2 How does AI work? - The Process of Machine Learning

The previous section has established that intelligence, be it human or artificial, is mainly characterized by acquiring knowledge from examples and experiences. In Artificial Intelligence, this is most recently, achieved to a high level through Machine Learning algorithms.

Machine learning is a generic term for the “artificial” generation of knowledge from experience, usually in the form of data. An artificial AI system learns from examples and can generalize these after the learning phase is complete. To do this, Machine Learning algorithms build a statistical model based on training data. This means that it does not simply learn the examples by heart but recognizes patterns and regularities in the learning data. It is applied in many different domains, for example, to detect credit fraud, to analyze stock markets or to recognize speech and interpret images [38].

The term “Machine Learning” should be distinguished from the term “Artificial Intelligence” and from the term “Deep Learning”. These three terms can not be used as synonyms as they each represent a subset of the other. Machine Learning is an integral part of AI applications today. In other words, Machine Learning is a subset of AI, which includes models such as Neural Networks, Support Vector Machines or Random Forests.

However, not all AI applications are Machine Learning applications. An example of this is a chatbot that relies only on a large data set made by humans to deliver an answer. Deep Learning, on the other hand, is a subset of Machine Learning, which will be further investigated in section 4.2.3. An example of a Deep Learning application is the interpretation of handwriting [39].

This section will take a more detailed look into the different subtypes of Machine Learning and what is required for effective Machine Learning.

4.2.1 Machine Learning Types

Machine Learning can be distinguished into three subtypes: Supervised learning, unsupervised learning and reinforcement learning [40]. Figure 3 provides a highly simplified over these three types and their distinctions.

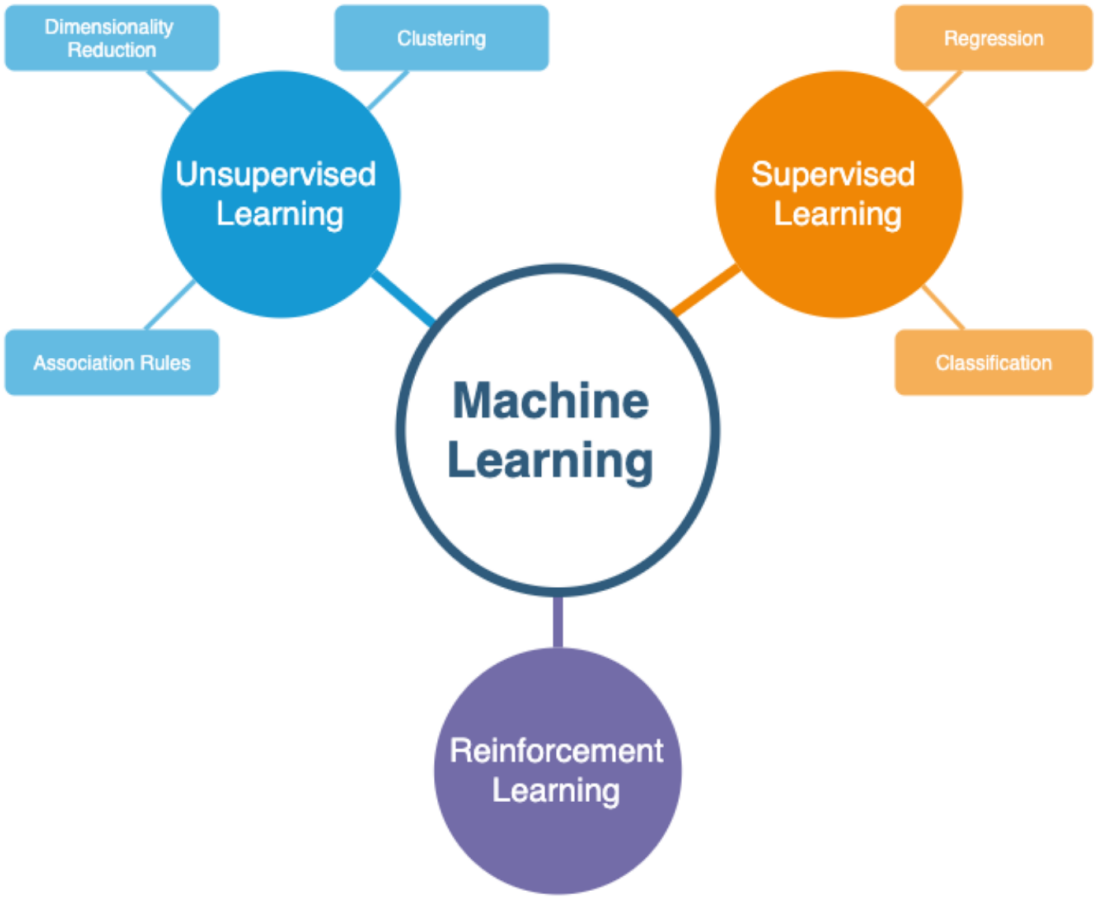


Figure 3: Overview of different Machine Learning subtypes and their distinctions [40]

Supervised Learning

Supervised learning is the most widely used type of Machine Learning [41]. Supervised learning algorithms build mapping functions based on labeled training data. The mapping function is then tested by calculating the output for each input and comparing it to the expected output (ergo, the label). If there is an error, the parameters in the function are adapted until there is no improvement anymore [42][43]. Figure 4 displays this process in a high-level visualization. Supervised learning can be further split into Regression and Classification.

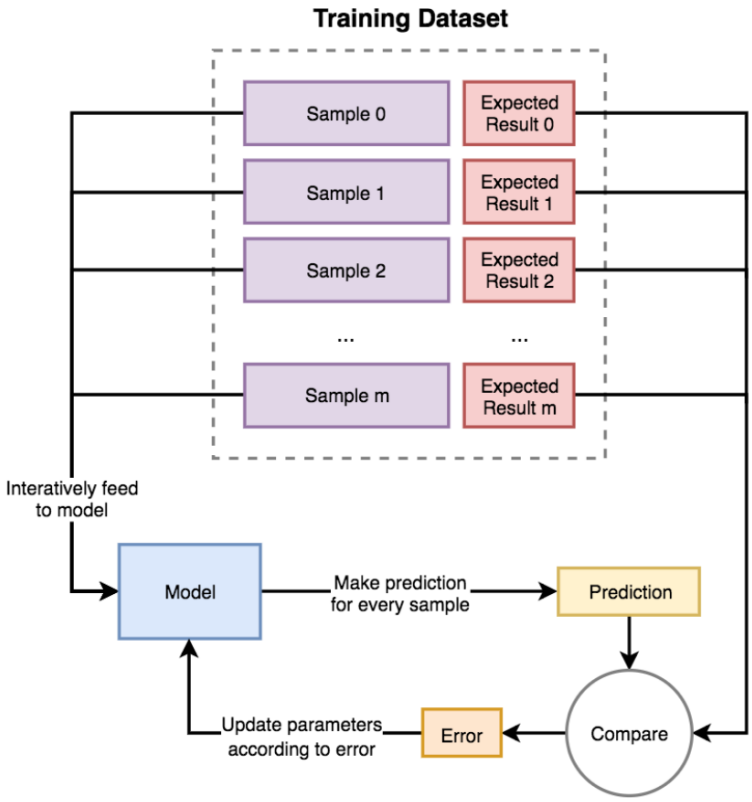


Figure 4: Schematic illustration of supervised learning [42]

Regression analysis is a statistical method that provides a connection between an independent and a dependent variable. There are several regression models with various amounts of variables and different mathematical models. However, all have the same purpose of analyzing the effect of the independent variable on dependent variables [44].

Classification algorithms assign a given input to a predefined class. For example, it can assign an incoming email to the "Spam" or "Not spam" class. Given that all possible outputs (classes) have to be defined before the algorithm can be trained, it is crucial to understand that the algorithm will assign the input to one of the classes, even if it does not fit into any of them [45].

Unsupervised Learning

Unsupervised learning is a method of Machine Learning in which the algorithm learns to recognize patterns and relationships in data in an exploratory manner independently and without supervision. The input data is unlabeled and has no predefined desired output. Figure 5 presents a schematic illustration of unsupervised learning's concept. The two most substantial areas of unsupervised learning are reducing data dimensionality and data clustering[42][46].

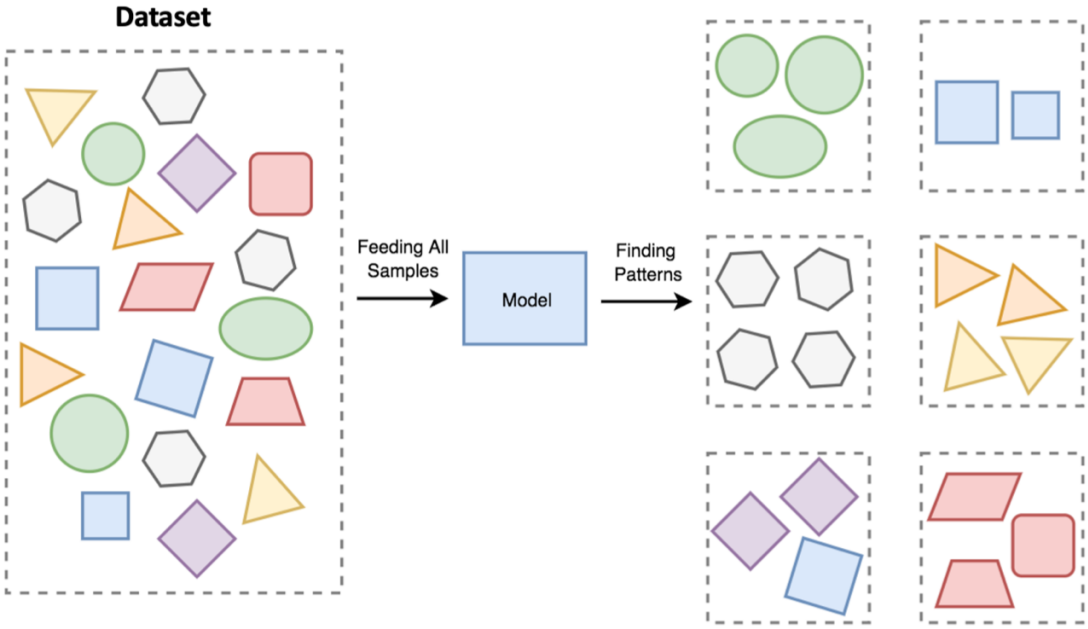


Figure 5: Schematic illustration of unsupervised learning [42]

For Pattern Recognition and Data Clustering, unstructured data with different feature values is fed into the model. The model then finds patterns within the data and clusters the data points into groups [47]. Figure 5 provides an illustrated example of this.

Dimensionality reduction limits the selection of variables present in the data to the essential and target variables. This method prevents the algorithm from learning only the specific patterns of the training data set (overfitting) and subsequently being unable to make any accurate statements about other data sets. Through this simplification of the model, the computational time decreases by reducing the numbers of calculations for classification because there are fewer dimensions to choose from [48].

Another application of unsupervised learning are association rules. This domain is concerned with finding strong rules in the data set that describe correlations between features. E-Commerce basket analyses are a prominent example of this [49].

Reinforcement Learning

Reinforcement learning deals with closed-loop problems in which the system's actions influence the environment which sends different inputs back to the system (agent) depending on the action, see Figure 6. Instead, the learner is left on their own to figure out which particular actions result in the best rewards, using a trial and error method [50].

Three features distinguish reinforcement learning from the other two learning types: Closed-loop problems, lack of direct instructions and lack of information about desirable actions [50].

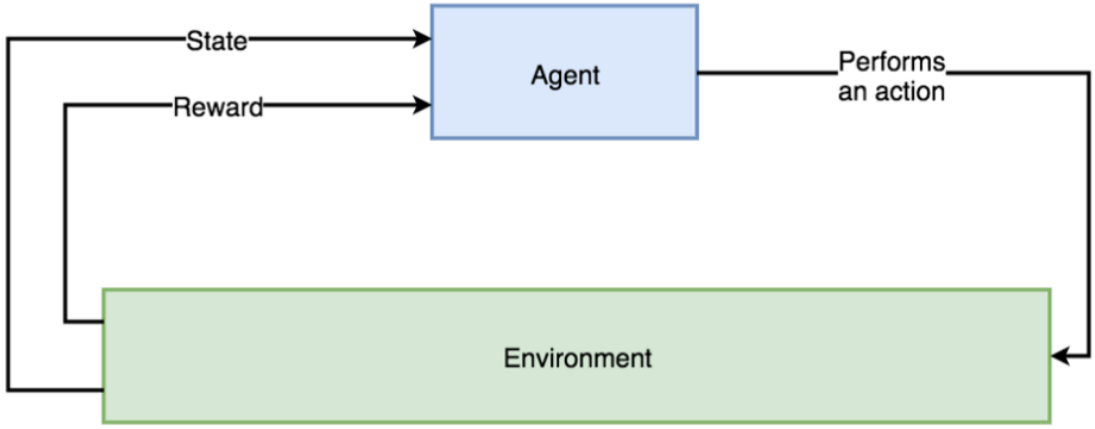


Figure 6: Schematic illustration of reinforcement learning [50]

In this type of learning, a goal-seeking agent is allowed to interact with an uncertain environment by choosing actions that influence this environment. In cases where planning is essential, reinforcement learning requires a balance between real-time action and planning selection to improve the environment model. One of the most important features of this type of learning is that it affects the future state of an environment. Reinforcement learning finds its applications in neuroscience, psychology and engineering, among others [51]

4.2.2 Digital Gold - Data Processing & Feature Engineering

All of the subtypes described in the previous section execute the same high-level task of transforming an input into an output. Figure 7 depicts the high-level process of training a model using training data, which is then used to calculate the outputs for "serving data" [41].

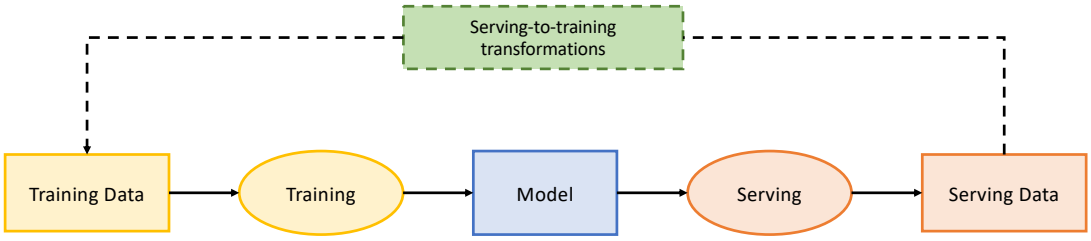


Figure 7: Schematic illustration of a general Machine Learning training process[41]

As data is always the input for such Machine Learning models, these models can only ever be as good as the data that is used to train them. Specifically, this means that training sets are large-scale data sets from a great variety of sources and that the data is clean [52].

The first requirement has become easier to meet in the past decades. Since the mid-2000s, computational systems, especially smartphones and IoT devices, have become very efficient at gathering and transporting vast amounts of data. This has resulted in high quantities of unlabeled data available for training Machine Learning algorithms [41].

The second requirement is not as easily met. Because data today is created at such a high speed, its quality is often second-tier. Thus, it cannot be used for training "out of the box". Therefore, this vast amount of unlabeled data must first be cleaned and processed. Clean data will have no duplicate, incorrect or missing data points, and it must be consistent [52]. The process to reach such a state of clean data can be divided into three steps [53]:

1. **Understanding:** The raw data must first be analyzed in order to identify anomalies and outliers and finally encode data into features that the algorithm can interpret
2. **Validation:** This step is the most crucial for the algorithm's quality. It ensures that the data set has the expected features and that those features have the expected values.
3. **Cleaning:** This step acts on the findings from the previous steps by removing all anomalies, outliers and errors, and by completing or deleting incomplete data.

This section has shown that labeling and cleaning the vast amounts of unlabeled raw data is an important but tedious process. Technological developments in Machine Learning have brought forth Deep Learning, which can automate these processes, reducing the amount of human labor needed for cleaning data and training models.

4.2.3 Deep Learning - When Machines teach themselves

As established above, conventional Machine Learning techniques cannot process data in its raw form and need human intervention to process the data. Deep Learning solves this problem by being able to abstract the given data through multiple levels until a complex function can be learned. For classification problems, these higher and more abstract levels highlight the important aspects of the input while suppressing irrelevant variations, for example, the background in an image. This ability to identify the relevant features without human guidance makes processing large sets of unstructured raw data much more efficient than using "traditional" supervised learning like kNN [54]. Figure 8 illustrates the Deep Learning multilayer network schematically.

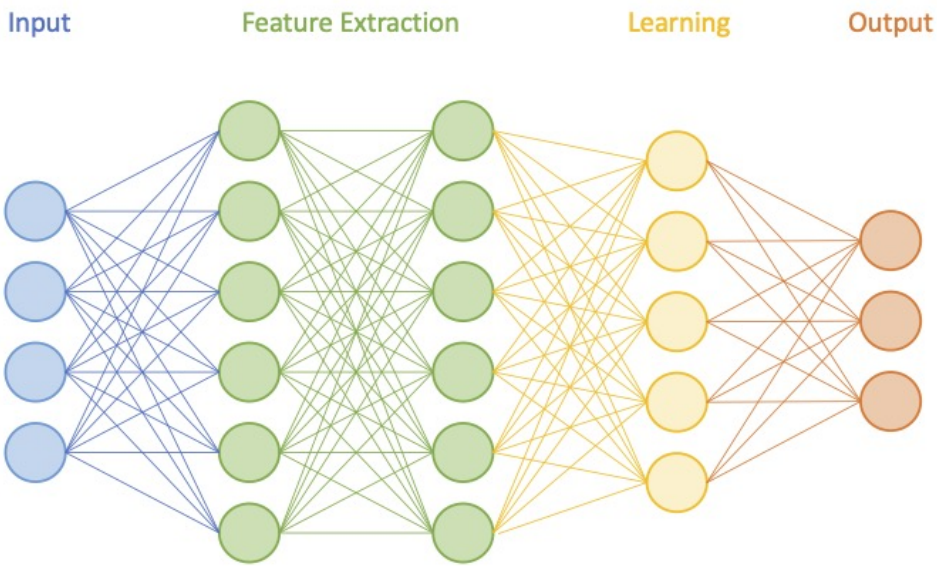


Figure 8: Schematic illustration of a Deep Learning multilayer network [55]

Deep Learning algorithms can extract features on their own because they learn to map an input to a given output by going from one layer to the next layer. The nodes of the layer calculate a weighted sum from the input they received and pass it through an activation function[2]. The output of that function is then sent to the nodes in the next layer. This means that the output of each layer is the input for the next layer until the output layer is reached, defining the result [41][54].

During training, the Deep Learning algorithms apply gradient-based optimization algorithms throughout the multilayered network to adjust parameters in each node based on the error of the result [41][54]. This method is called backpropagation. It starts at the

[2]An activation function is a non-linear transformation. An example is the ReLu-function, which outputs the value if its positive or 0 if the input is negative, or the Sigmoid-function which transforms the input into a value of 0 to 1 (often used to calculates probability [56])

output layer by calculating the error from the given output to the desired output. Then it will gradually go backward through the layers, adjusting the parameters of each layer. Then the task is repeated and if the error persists, backpropagation is applied again. This cycle repeats until the error rate of the training set is small enough [41][54].

Because it can work so well with large amounts of unstructured data, Deep Learning has led to significant improvements in computer vision and speech recognition [41]. The following section will focus on more specific examples of AI, Machine Learning and Deep Learning Use Cases.

4.3 Who does AI affect? - AI Use Cases

It has been mentioned often that AI finds application in almost every industry and aspect of life. This section provides specific examples of AI applications in different industries. The list is not exhaustive, but it proves that the statement of AI finding its way in all industries is indeed true.

Consumer Goods & Services. AI is used in many situations to interact with customers, for example, through intelligent chatbots on a website or through adverts being placed on a website by an algorithm [57]. AI is also used to cross- or upsell products through recommendations [54]. Some examples are the personalized movie library in Netflix or Amazon's "you might also like" section in their online shop.

Industrial Goods & Services. AI is not only present in the consumer domain. In the manufacturing industry, AI is used with great benefit through predictive maintenance. This allows businesses to determine the condition of a machine. Thus, the remaining useful life of that machine can be estimated and production downtime is eliminated by being able to replace the machine before its end-of-life [58].

Materials & Construction. In the construction industry, AI can predict cost overruns in construction projects and identify conflicts between different designs in a Building Information Model. A major benefit comes from identifying and pointing out safety hazards based on pictures from construction sites [59].

Energy & Utilities. Similar to other industries, AI can help the Energy and Utility industry to better use their resources by predicting usage and detecting defects in pipes, wiring or machines early so that they can be replaced without any downtime [60].

Financials & Real Estate. A very prominent example of AI application is fraud detection, where AI algorithms identify anomalies in financial transactions [61]. In Real Estate, AI is already helping investors and companies to find interesting building sites and it can also design a corresponding building plan for the site [62].

Health Care. The Health Care industry is another popular provider for AI use cases. For example, AI can identify tumors in images much earlier and more precisely than oncologists. There are also AI applications that can diagnose diseases from answers to a questionnaire [63].

Primary Sector (Agriculture). AI is also relevant in the primary sector as it can help farmers with crop management and, for example, more efficient distribution of agrochemicals [64].

Technology & Telecommunication. AI is at the core of today's most popular social media apps. It takes on the main tasks of filtering and distributing content to the users interested in it. Search engines, which hold an essential role in guiding users through the internet, also use AI to understand search prompts and prioritize results based on relevance [54]. Another interesting application is that AI can write software code [65].

These use cases show that AI can affect everyone in their professional and personal lives as it has already taken hold of all major industries.

This chapter has established that Artificial Intelligence is not a mere "copy" of human intelligence but a technology that has evolved significantly over the past few decades and is transforming many industries and aspects of our lives in major ways. However, as Artificial Intelligence is being more widely adopted, discussions about its downsides and challenges arise.

5 Ethics & Challenges

As many beneficial developments in Artificial Intelligence technology arise, so do questions about ethical implications and harmful AI applications. This chapter provides examples of such dangerous AI applications and categorizes them into five AI challenges to display why AI ethics is necessary. Furthermore, it analyzes which ethical principles for AI are in place today and why ethics in the context of self-regulation does not seem to be efficient enough to prevent harmful AI applications.

5.1 Why AI needs Ethics - The Limitations of AI and related Consequences

The following five categories of AI challenges provide an overview over the most pressing problems in AI applications. Such an overview enables businesses and policy makers to properly identify which areas need to be addressed in an effective AI regulation.

5.1.1 Opacity and Lack of Explainability

Algorithms are continuously being implemented in more and more areas where they significantly impact the lives of the affected parties (see examples in 5.1.6). As is the case with human decisions in such delicate situations, it is crucial to base algorithms' decisions on a cohesive and thorough explanation. But with opacity being an essential characteristic of self-learning algorithms, this proves to be a tough challenge. However, the model itself is not the only component where the explainability of an algorithm is failing. Opacity is also caused by poor structure in data and code and a lack of tools to visualize and track large volumes of code and data, which would help understand the drivers behind the decision [66].

Algorithms build their understanding of the world based on the data they receive. As algorithms' main task is to find patterns in data, they will always achieve this task, whether there are actual causal correlations between variables or not [66]. These random correlations can lead algorithms to make a decision that humans will not be able to understand or explain. In contrast, a more balanced and complete dataset could lead to an understandable decision that reflects the real world. Given this dependency, it is crucial to provide the algorithms with a most accurate representation of the real world, so they can make decisions that are conclusive in it. It is clear that representing the entire environment of a "real world" would require a volume of data that cannot be efficiently handled by current hardware. Abstraction in a responsible way is therefore necessary, but such abstraction needs to be explainable as well: Researchers and developers must be able

to explain why certain variables were chosen or omitted for a training data set [66].

It is also necessary to provide users with enough information about the features and limitations of an algorithm. This will enable them to understand, interpret, and evaluate its decisions adequately. But if too much or poorly structured information is provided, it can overwhelm users and render the algorithm more opaque than it is. At the same time, many companies do not want to disclose their source code out of fear of exploitation because individuals who understand the code could influence the algorithm to produce a particular result by tweaking their input [66].

The failure to focus on explainability or transparency of algorithmic models and training data can compromise the justifiability of algorithms' decisions or propositions. It can even lead to a lack of scrutiny or accountability (see chapter 5.1.5).

5.1.2 Discrimination and Bias

As established in the chapter above, the quality of underlying data has a direct impact on the quality of the algorithm's conclusions. Decisions or calculations are only as diverse and as fair as the data they are based upon. The problem is that existing data points often reflect history and therefore existing social bias. If such data points build the basis for training algorithms, these algorithms will naturally reflect this social bias. A similar result is achieved when data is consciously or unconsciously sampled preferentially, for example, in favor of a certain social group. Finally, bias also can be created when an algorithm is applied in an environment it was not designed for, for example in another country with other cultures and values. Bias is therefore not something that is created only in the design phase of the algorithm. In fact, it can occur at any stage of the design, development or deployment process [66].

Although researchers do not yet agree on a definition of algorithmic fairness, it is evident that such discriminatory algorithms must be prohibited. The obvious approach to reduce bias in algorithms is to take a proper look at the training data to identify possible bias and reduce it through the data cleaning & transformation process, for example leaving out protected variables like race or gender. However, social contexts must be taken into consideration as such protected variables can sometimes be appropriate. For example, women's reoffending rates are significantly lower than men's. Excluding the gender attribute in such cases would lead to women getting ranked disproportionally high by an algorithm. Other theories see it beneficial to use artificially generated data to diversify training sets or to consciously incorporate conflicting bias to counter out the original one [66].

If no measures against bias are implemented, such biased algorithms can then lead to

discriminatory actions suggested or taken by the algorithm, disproportionally affecting groups that are already disadvantaged in society [66].

5.1.3 Manipulation

Manipulation occurs where users' autonomy and freedom to decide is undermined. Especially in the age of social media with selective content distribution, this issue is more present than ever. These social networks base on algorithmic profiling, which tracks user profiles and profile groups. Such profilings can impend a user's freedom to think, communicate and form relationships because the algorithms only make a selection of information available to them and/or prevent them from easily accessing information of different perspectives [66]. Another aspect of manipulation is the deceiving of people through altered or fake content. Generative AI applications can be abused to create fake images or text, for example the GPT3 algorithm or Deepfakes [67][68]. Notable examples of manipulation through algorithms include Facebook influencing the vote in the United States presidential election in 2016 [69] and disinformation spreads during Russia's attack on Ukraine [70].

Attempts to overcome this challenge include constraining the autonomy of algorithms and being able to reverse their actions [66].

Failure to address this challenge would lead to a grave violation of fundamental rights to autonomy which can only have negative effects for the public's trust in and therefore the future of AI.

5.1.4 Data Security

Every day, about 2.5 quintillion[3] bytes of data are created through Social Media, IoT devices and other means [71]. This makes it easier than ever to gather input for AI algorithms. The means through which consent to use this data is asked for remains ethically questionable: Oftentimes apps that gather such data are free, but users "pay" the company with their personal data [72]. This data is then used to train algorithms. The users giving their data in most cases have no idea what algorithms are trained with their data or what goals these algorithms pursue. A similar problem arises when such input data is processed by algorithms without the user knowing or consenting to it.

Another security and privacy issue is that algorithms are code, which process data, and code is susceptible to hacking. Cyberattacks have grown in the past years and more than 10'000 cyberattacks have been reported by Swiss companies in the first half of 2021 alone

[3] 10^{18}

[73]. AI algorithms pose no exception, so it should be imperative to protect them and the data they process from such threats as best as possible.

To give web users more authority over their own data, the EU has established the general data protection regulation GDPR[74]. This regulation clearly defines scenarios in which companies are allowed to process user data. It also requires measures to protect the data stored by companies from outsider threats like hacking. While such regulation can mitigate the privacy and security challenges, it is only applicable in EU jurisdiction. Switzerland has revised its data protection law to align with GDPR in many cases but the law hasn't yet been enacted [75]. China has also put into effect its version of GPDR called Personal Information Privacy Law (PIPL), while the US has no federal data privacy law at all [76].

While GDPR seems like a model that could be adapted by many other countries, security & privacy still remains an issue for AI. The law only covers the security of data, not the security of the algorithm itself. Also, algorithms pose new threats to data privacy, for example through the issue of mass surveillance where facial recognition technology can be used to process security camera footage.

If security and privacy issues of AI are not (fully) addressed, people's trust in AI technology might be impacted negatively as it is an invasion of privacy to use people's data without their consent.

5.1.5 Liability

In hybrid systems where artificial and human agents work together to achieve tasks, it is difficult to trace back responsibility for the actions performed. An especially important aspect in this regard is the level of trust the human agent has in the artificial agent. Because of their supposedly scientific nature, algorithms could lead even experts to avoid questioning an odd suggestion because "the computer said so". This offers a possibility for companies to distance themselves from any liability by "blaming the algorithm". Through the problem of "black-box algorithms" (see chapter 5.1.1) this issue is intensified as companies can never fully understand why algorithms came to a certain conclusion [66]. The question then is, how can they be held accountable for its actions if they didn't exactly program it to do them?

Some researchers suggest that all organisations should take responsibility for their algorithms regardless of how opaque they are. Others would hold all agents morally responsible that were involved in the situation in question [66]. Both these approaches lead to the assumption that responsibility can never fully be given to AI algorithms.

As long as the question of liability remains unanswered, there is no legal ground to appeal decisions made or influenced by algorithms. This makes it hard for affected parties to take

action against wrongful behavior towards them.

5.1.6 Examples

Having introduced the five categories of AI harm, this section will describe specific situations where Artificial Intelligence algorithms caused unnecessary harm.

Algorithms classify Swiss prisoners

Since late 2018 all German-speaking cantons in Switzerland apply risk-oriented sanctions (ROS) enforcement. Part of this risk-based approach is to calculate the risk of recidivism of each convicted criminal. This is done by an algorithm called "Fast Screening Tool" (FaST), which classifies prisoners into three risk categories based on data from their current case as well as their criminal history. The data of prisoners in the highest-risk area gets analyzed again by an algorithm called Fotres, which proposes intervention methods or therapy suggestions for the inmates during their prison time [77]. Figure 9 displays the two algorithms and their purpose involved in ROS.

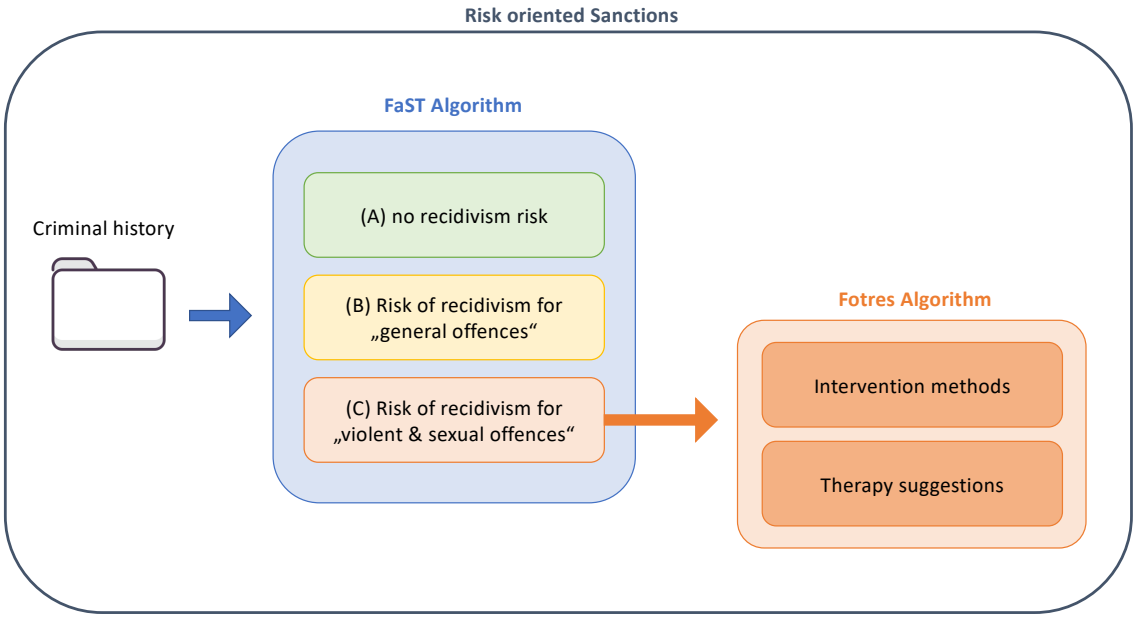


Figure 9: Graphical representation of the algorithms involved in ROS.

Justice institutions like District Attorney's Offices as well as courts and prisons trust the system to indicate whether a criminal should be prosecuted and whether they receive sentence mitigation. While justice institutions endorse the use of the algorithm through professionals, insiders and researchers have issued concern about the effectiveness and scientific basis of the system [78][79][80]. The algorithm has only been scientifically examined twice - one of which was performed by Fotres itself. The two evaluations have come to differing results, where the internal evaluation supposedly proved the effectiveness

of the system, the external study conducted by scientists at the University of Ulm found that Fotres may be beneficial but also carries the risk of "Pseudo Security" [79][80]. Others, for example the cantonal judge of Lucerne, have criticized the fact that the evaluation bases solely on data of previous and current cases. They believe that before a decision with such an impact is made, the surveyors must be in direct contact with those affected - in order to get a picture for themselves [77].
The fact that prisoners are not allowed to weigh in on the decision the algorithm has made and that they have little means to appeal the result, which will directly or indirectly impact their sentence, along with the questionable scientific basis of the algorithm raises concern. There is certainly more evaluation needed in order to decide whether the algorithm indeed surpasses the estimates made by professionals in the justice system.

Health Care Algorithms exhibit Racial Bias
Algorithms in health care are becoming more common, for example to predict illnesses or analyze medical images. But certain health algorithms, for example the heart disease prediction algorithm CKD-EPI, are under suspicion that they are racially biased[81][82]. These algorithms differentiate between "Black" and "non-Black" skin tones, where patients classified as having Black skin are less likely to receive a correct prediction, which could lead to them missing out on necessary treatment. According to researchers, this differentiation between Black and non-Black skin tones is necessary because the algorithm to predict the risk of kidney stones is based on creatinine levels. Black people are assumed to produce more creatinine, because of generally more muscle mass. Therefore, the algorithm corrects the calculation for Black people in general. However, studies have cast doubts on whether this claim is based on scientific research [82]. Researchers and activist groups like Algorithm Watch express concern about these algorithms because they use a social marker instead of genetic or individual properties of the patient. Several studies published in the Journal of the American Medical Association between 2019 and 2020 have found that such algorithms indeed lead to poorer results for Black patients [83][84][85]. This has strengthened the calls for discontinuation of such algorithms [86]. While this led many US hospitals to prohibit the use of such algorithms, they are still among the most popular algorithms in Switzerland. The CKD-EPI algorithm for example is used in all five Swiss university hospitals [81].

This example displays that self-regulation reaches its limits when it comes to trust in an algorithm. While an algorithm can be prohibited in certain institutions, others apply exactly the same algorithm most frequently. This paradox can lead to mistrust. A regulatory framework could set a clear bar for what requirements must be fulfilled in order for an algorithm to be "safe". That way stricter handling is enabled, but people using algorithms and those that are affected by them can still be sure that it remains safe.

Facial Recognition in Law Enforcement

The cantonal police Aargau is using a facial recognition system called "Better Tomorrow". The system is used two to three times a week by a specialist to compare photos against the cantonal and inter-cantonal database of faces. The system then delivers a collection of "possible matches", of which the specialist decides which ones are followed up upon. According to the head of the cantonal police, the algorithm generates 10-15 percent more clues on perpetrators. However, the accuracy of the algorithm cannot be evaluated since the department that uses the algorithm gets no feedback from the department that interrogates and arrests the suspects [87]. This is concerning, as facial recognition is far away from perfectly accurate and assuming it is without evaluation can have major implications for certain ethnic groups and genders, especially women of color[4] (see figure 10).

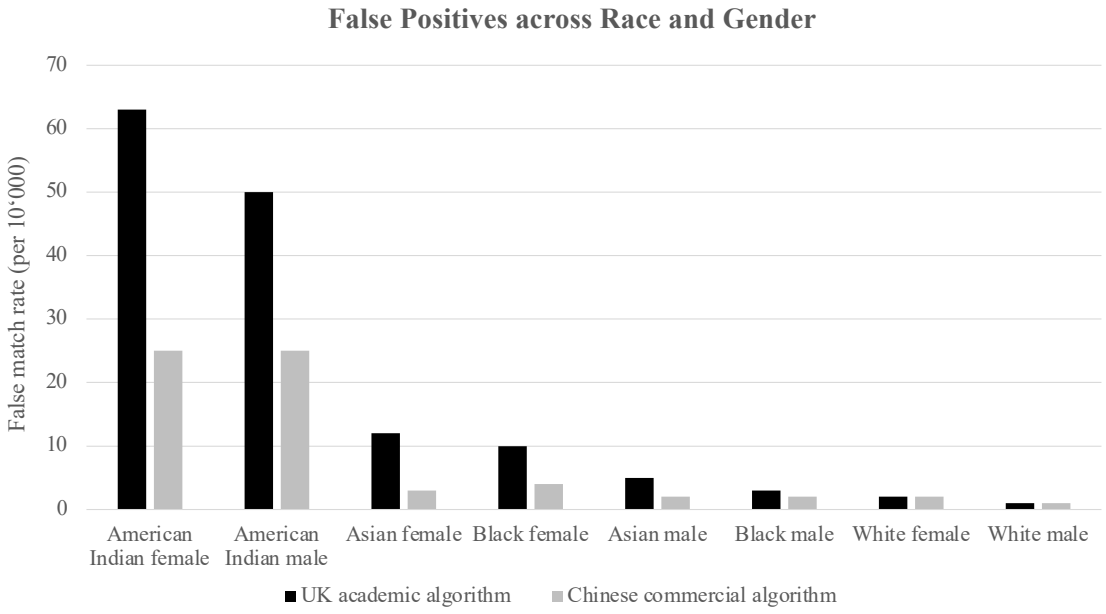


Figure 10: Review of facial recognition algorithms displays that false positive rate is highest for female people of color (shown with two examples) [88].

This is illustrated by the application of facial recognition in law enforcement in other countries. Misidentifications have put people in the situation of having to prove that they are not the person the algorithm says they are. In January 2020, a black man was arrested by Detroit police because the facial recognition had matched the image of a thief to his driver's license. He was detained for 30 hours, although a human could have identified that he and the target were not looking alike. In August 2021, London Police announced they would integrate facial recognition technology despite their previous experiments in 2020 having resulted in only 8 out of 42 matches being correct. Although the officers

[4]In this paper, the term „People of Color" is used when a certain problem affects multiple ethnic groups in general (excluding Caucasians) rather than just one specific ethnicity.

recognized 16 matches as non-credible, experts say that humans often confirm a system's bias rather can correct it [88].

A lack of regulation can let law enforcement implement algorithms that base on a technology that is not yet unbiased. Even major technology companies like Amazon, Microsoft and IBM have declared to not sell their facial recognition technology to law enforcement until there is a strong legal basis [1][2].

Facial Recognition and Mass Surveillance in Swiss Patrol Cars

Another example of problematic Artificial Intelligence applications in law enforcement refers to a unique patrol car model by Ekin Technology. The Turkish company designed a patrol car that can measure speed, identify license plates and identify people in proximity using facial recognition. It is already used in the USA, Abu Dhabi and Azerbaidschan - and some Swiss police forces have tested it for use in Switzerland [89].

The car is equipped with sensors and 360° surveillance cameras. The footage is automatically run through facial recognition software that compares the identified faces to a list of people of interest [89]. This functionality would allow for constant mass surveillance and identification of people without their consent.

In 2019, the vehicle was still in the process of being evaluated for its functionality to "read" license plates but no evaluation of its sensors and facial recognition technology was planned. Without a legal basis for facial recognition technology, such vehicles - or technologies, respectively - could be allowed in Switzerland without any background checks [89]. The lack of a legal basis for AI in law enforcement allow an unsafe and untested facial recognition technology to be applied in a domain where wrong decisions will significantly impact the lives of the affected parties. In addition, it can lead to mass surveillance, which is critical in its own right and undermines people's right to privacy.

Misidentifying People or Characters with Black Skin in Pictures

The accuracy of facial recognition algorithms has not only been questioned when applied in law enforcement but also when identifying people in photos or labeling pictures. In September 2021, Facebook's[5] facial recognition algorithm sparked controversy when it identified two Black men as "primates" (see screenshot in figure 11) [90].

[5]During the writing of this thesis, the company has been renamed to "Meta". Because the issue described arose when the company was still called "Facebook", this name will be used to refer to the company in the context of this study for clarity reasons.

Figure 11: Facebook Algorithm identifying Black men as "primates" [91]

Two months later, Facebook announced it would stop using facial recognition software and delete all facial data in its database, explaining that they weighed the many benefits of facial recognition technology against the growing concern about the technology. In the announcement, they did not specifically mention the incident of the algorithm labeling the Black people as primates [92]. But Facebook is not the only Tech giant having dealt with problems in its computer vision / facial recognition algorithms. An experiment conducted by AlgorithmWatch has revealed that Google's Computer Vision API labels pictures of Black comic characters as "mammals" while this label is missing in pictures of the same character with white skin. A Google Image Search revealed that a search with the term "mammal" results in only pictures of animals, not humans. A similar issue was announced on Twitter when a user found that the iPhone labeled pictures of Black comic characters as "animal" whereas it didn't attach this label to the same picture of characters with white skin. However, this experiment has neither been scientifically conducted nor could it be reproduced with the iPhone of this study's author [93]. Nevertheless, the commonalities in all three algorithms indicate that - at least in some cases - they have trouble identifying and labeling pictures of dark-skinned people, whereas these problems do not arise with pictures of white people.

At the moment, companies seem to only adapt these algorithms or refrain from using them when they are under public pressure. A regulatory framework should force them towards catching and removing such imbalances in the training phase, before any harm has been done, instead of after deployment when biased algorithms can cause serious harm.

Problematic Algorithms for Social Welfare Programs in Europe

Many countries are taking up AI in public administration (see chapter 7). However, these algorithms are not always working as expected as the following examples illustrate.

In August 2021, it has come to light that an algorithm used in France to calculate social welfare payments has automatically and wrongly sent out debt statements to eligible applicants. For a woman working a regular job as well as a freelance job, the case seemingly got "too complex" for the algorithm, according to a social worker using the algorithm. Subsequently, it automatically sent out a debt statement of over 90% of her regular payments. The error could only be identified because the woman directly contacted her social worker. Although the case has promptly been corrected, it raises concern as according to French law, decisions made by an algorithm are void if the person affected has not been informed about the involvement of the algorithm. The woman states that she was at no point informed about the use of an algorithm [94]. This leads to the suspicion that had she not intervened, the debt would have been taken from her monthly payments without further consideration.

Two years earlier, in 2019, two similar cases happened in Sweden and in Spain. In Sweden, an algorithm had wrongfully declined several thousand of applications for social payments. The algorithm was supposed to check whether the applicant fulfills the requirements and to send out warnings and hold back payments if this is not the case [95]. In Spain, an algorithm was supposed to decide whether an applicant was allowed to receive electricity contribution payments from the government. Within an experiment, a non-profit organization found that the algorithm declined certain profiles (i.e. widowed retirees), although they would be eligible according to the requirements [96].

These cases indicate that algorithms in public administrations, especially in the social welfare field, can have a great impact. The people in need are often the ones where a little error like the cut of a social welfare payment can have major implications on that person's life, as they already do not have much money.

Biased and Opaque Hiring Algorithms

In the USA, more than 100 employers apply a hiring algorithm called HireVue in their recruitment process [97]. The algorithm analyzes video interview footage of applicants and calculates an "employability score" from several thousand parameters like word choice and facial movements. With these employability scores, HireVue ranks candidates into

high, medium and low tiers based on their "likelihood of success". After this, a human HR employee chooses which participants to pass on to the next stage in the recruitment process. HireVue claims to be a more objective party in the hiring process contrary to human recruiters. After a finished interview, HireVue neither discloses the score to the candidate, nor lists where they did something wrong. Since the algorithm is looking for the perfect candidate based on a description of current and past employees that do well on the job, it can only be assumed that it looks for someone that is very similar to the people that are already holding the position. This could end up penalizing non-native speakers, nervous interviewees or anyone else that does not look and speak the same as the existing team. For example, a woman might be penalized when applying for a job that is mostly done by men because she speaks and gestures differently, even though her qualifications are the same. Although there is always a human making the algorithm's decision in the end, scientists say that systems like HireVue can seem convincing due to the mass of data points they create, even when the method of creating this data is not backed by science. Therefore, although a human might have the possibility to select someone from the lowest tier to pass to the next phase, they might not choose to do so because the algorithm, which they have much confidence in, has already delivered them enough prime candidates in the first tier [97]. An article by ElleXX.ch suggests that such algorithms are also applied in Swiss companies, but this could not be verified by other sources [98].

The example of HireVue negates the assumption that an algorithm can not do harm when the final decision is made by a human. When an algorithm seems convincing and delivers convincing data points, its scientific basis and unbiased nature are not always questioned. Therefore, humans over-confidently trust the algorithm and do not look into the pile of applications that have been labeled the "lowest tier". It also indicates that algorithms can be deployed, commercialized and rise in popularity without a proper evaluation. When questions arise after this rise in popularity, there is already harm done. With a proper regulatory framework, such harm could probably be avoided when applications are evaluated properly and in a uniform way internally or externally.

Russian Troll Farms use AI to spread Disinformation about Ukraine-conflict

During the 2022 Russia-Ukraine conflict, Russian Troll Farms used AI tools like ThisPersonDoesNotExist.com [6] to make fake user profiles on social media appear more believable (see figure 12). These profiles then spread disinformation, such as Russia's operation going better than it was in reality [68].

[6] A website where an AI generates realistic images of people's faces

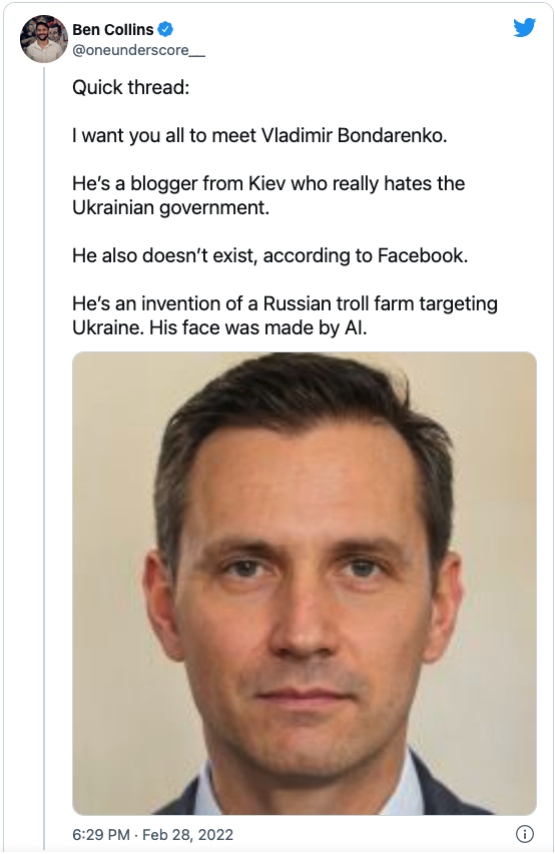


Figure 12: A Twitter user identifying the profile picture of a fake user spreading disinformation [68]

A few days after these profiles were discovered, a deepfake video was displayed on a hacked Ukranian TV station. The video showed the Ukranian president Zelenskiy asking Ukranians to "put down their arms" [99].

Although Social Media platforms like Facebook and Youtube took these posts and profiles down after identifying them as fake, they still struggle to differentiate between real and fake news. For example, a post by a French politician about the conflict was falsely identified as fake news and taken down as well. On the other side, their algorithms helped spread the disinformation to the point where posts of these profiles received hundreds of thousands of views and clicks [70]. This is just one example where AI plays an essential role in spreading disinformation. It can provide deep fakes as profile images or video material to manipulate the narrative and automate the writing of fake news articles. A test by MIT's Technology Review has shown that OpenAI's GPT3 algorithm can write a coherent short fake news article with only a few words of input [67]. This fake news article is displayed in figure 13.

Russia has declared war on the United States after Donald Trump accidentally fired a missile in the air.

Russia said it had "identified the missile's trajectory and will take necessary measures to ensure the security of the Russian population and the country's strategic nuclear forces." The White House said it was "extremely concerned by the Russian violation" of a treaty banning intermediate-range ballistic missiles.

The US and Russia have had an uneasy relationship since 2014, when Moscow annexed Ukraine's Crimea region and backed separatists in eastern Ukraine.

Figure 13: Fake news article generated by OpenAI's GPT3 [67]

5.1.7 Summary

The five challenges manifest that AI technology poses many challenges within the entire life cycle of an AI application: From gathering data and cleaning data sets to the question of who is held responsible for harm caused by algorithms. Table 2 shows that addressing the challenges of AI with an AI regulation framework could have prevented many of the examples above.

Challenge	**Description**	**Examples**
Explainability	Opacity is an essential characteristic of self-learning algorithms. Poor structure in data & code can reinforce this opacity. Such opacity and the resulting inability to explain algorithmic decisions limits the users' ability to understand, interpret and evaluate such decisions adequately. This will compromise the justifiability of algorithmic decisions resulting in a loss of trust and faith in the legitimacy of the technology.	Social Welfare Algorithms that came to unexplainable conclusions like disapproving rent incentives, HireVue not disclosing its rating algorithm to candidates, Fotres being questioned for its scientific basis

Table 2: Overview of ethical challenges for AI

Table 2 (continued)

Challenge	**Description**	**Examples**
Discrimination/ Bias	Algorithmic decisions are only as diverse and as fair as the data they are based upon. But oftentimes, existing social bias is embedded into datasets used to train algorithms. This leads to algorithms reinforcing the bias by negatively affecting groups that are already disadvantaged in society.	Facial Recognition & Computer Vision software not recognizing faces of people of color with the same accuracy as white faces, sometimes even mistaking them with animals HireVue granting lower scores to candidates who differ in characteristics from previous position holders
Manipulation	Selective content filtered through algorithmic profiling makes it hard to gain a complete picture of a situation from different perspectives. This impends a user's freedom to think, communicate and form relationships. AI can also help automate the production and spread of "fake news" by creating deceptively real looking pictures or footage or by producing coherent fake articles. Such impediments of a fundamental right will certainly foster mistrust of AI applications amongst the public.	Russian Troll Farms using deepfakes to spread misinformation
(Data) Security	Data is gathered for and processed by algorithms often without users knowing what exactly their data is used for. Also, algorithms are code and therefore susceptible to hacking. Such security and privacy impediments through AI could gravely impact society's trust in the technology.	The Ekin Patrol car capturing and identifying each face around it within seconds, leading to mass surveillance.

Table 2 (continued)

Challenge	**Description**	**Examples**
Liability	Because of their seemingly statistical nature, algorithms can create a feeling of false security. People might stop questioning algorithms' decisions because "the computer said so, so it must be true". This creates opportunities for companies to distance themselves from any liability by blaming the algorithm. Liability mustn't be delegated to an algorithm, as this would create a scenario where people would lose the ability to appeal wrongful algorithmic decisions.	Social welfare algorithms in Europe were allowed to make decisions and act upon them that were later identified as wrongful decisions. It was only possible to correct these faults because humans noticed the errors and humans were involved in other stages of the decision process.

5.2 What is AI Ethics? - Ethical Principles

In order to overcome the challenges described above, many researchers and companies have formulated ethical guidelines. These guidelines can be summarized in five principles, which build a common ground to understand what ethical AI should be [100]. Table 3 describes these principles and which challenge they attempt to tackle.

Principle	**Description**	**Challenge targeted**
Beneficence	AI must be beneficial to humanity. This includes respect for fundamental rights as well as the sustainable development of algorithms.	General
Non-Maleficence	AI must be robust and secure. It mustn't infringe privacy or do any harm.	Manipulation, (Data) Security, Liability
Autonomy	AI must protect and enhance human autonomy as well as the ability to choose between alternatives.	Manipulation
Justice	AI must be fair, diverse and inclusive so it can benefit as many people as possible.	Bias & Discrimination
Explicability	AI systems must be understandable, transparent and explainable.	Opacity

Table 3: Principles for ethical AI (in style of [100])

5.3 Why Ethics alone doesn't work - The Role of Self-Regulation

Although many companies involved with AI algorithms have started to define ethical guidelines corresponding to the principles above, it remains questionable to what extent the principles are actually implemented. McNamara et al. even found that such measures are almost completely ineffective and do not change the behavior of professionals from the tech community [101]. Examples like the ones mentioned above prove this thesis right and bring to the forefront that the problem lies in the implementation of guidelines. This is due to two major issues.

5.3.1 Reason behind Self-Regulation

Self-regulation is sometimes used by companies and industries to mitigate or pre-empt truly binding government regulation (see chapter 6). Whenever the call for commitment to legal regulation arises, companies can highlight their membership in industry associations or their own ethical guidelines to stifle that call. In doing so, accountability is seemingly transferred from state authorities to the company or industry itself [72]. But companies can not only influence governmental regulation with self-regulation but also calm public opinion by using ethical guidelines as a PR instrument [102, p. 56].

If self-regulation is only put into effect with the sole reason to stifle or influence other forms of regulation and without any moral background, an ethical guideline alone won't help in creating more ethical AI.

5.3.2 Lack of actionable Recommendations

For companies that have implemented self-regulation for truly moral reasons, another problem arises: Ethical principles do not equal ethical practices[72][100]. There is a lack of concrete actionable recommendations that help implement the principles in the algorithms. Not implementing such practices can heavily impact the future of AI by undermining public opinion or reducing the adoption of algorithmic systems altogether [100]. Morley et al. have therefore created a requirements catalog that caters to the principles described above [100]. Such a catalog can help define tangible practices for designers and developers to adopt when creating AI algorithms.

Further potential lies in the education about ethics in algorithms. Engineers and developers are rarely educated about ethical issues nor empowered to raise ethical concerns. Therefore, even if issues are discovered, they are not discussed and result in the final product having little to do with the ethical guidelines [72].

This section has displayed that incorporating ethics into the AI application life cycle is necessary to overcome the challenges the technology poses. Although many companies have defined ethical principles to address such challenges, they fail to be effective. This confirms that self-regulation itself is not enough to address and overcome AI challenges. Because of this, the next section will explore the different types of regulation and what needs to be taken into account when regulating new technologies.

6 Regulation

Regulators often find themselves in a position to decide *what, when,* and *how* to regulate. What to regulate is already agreed to be Artificial Intelligence technology based on the topic of this thesis. This chapter presents seven types of regulation that cater to the question of how to regulate and discusses when to regulate from the standpoint that technology innovation cycles are much shorter than they used to be.

6.1 Definition of Regulation

Researchers agree that there is no one true definition of the term "regulation" [103]. To further analyze the topic of AI regulation, however, it is necessary to have a clear understanding of what is seen as "regulation". In this thesis, the definition of the term is based on the Cambridge dictionary. It sees regulation as "the rules or systems that are used by a person or organization to control an activity or process, or the action of controlling the activity or process" [104].

The following section 6.2 further describes the definition by explaining who or what can be seen as "person or organization" and how these parties can enact the "control" of an activity.

6.2 How to regulate? - The seven Types of Regulation

According to Steurer, three actors can enact regulation: Government, Businesses, and Civil Society [105]. They can enact regulation themselves or collaborate with one or both of the other actors. This results in seven types of regulation, which can then be split further by looking at how binding the rules are. The following figure 14 provides an overview of the seven types and their respective subtypes.

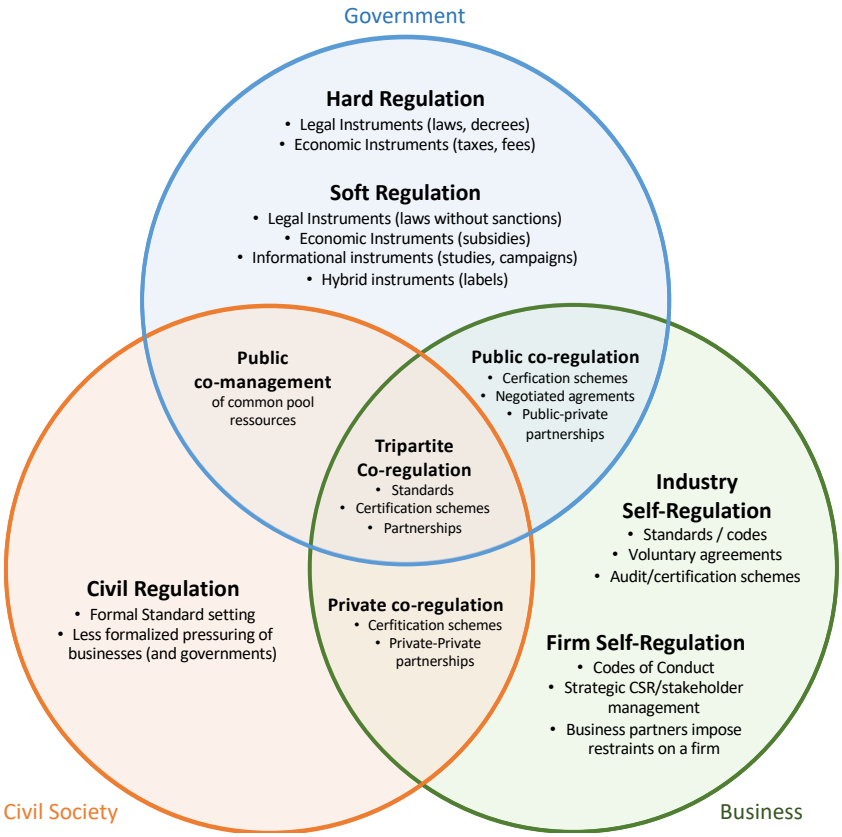


Figure 14: Overview of Regulation Types [105]

6.2.1 Governmental Regulation

Regulation by the government is possibly the most binding form of regulation that impacts all actors who interact with the regulated object. This type of regulation uses laws and economic instruments to drive market forces in the desired direction.
The subtype "hard governmental regulation" is binding for everyone because the executive and judicial branches ensure compliance with the rules. Typical instruments of hard regulation include laws, decrees, and obligatory economic instruments like taxes or fees. On the other hand, soft governmental regulation is more suggestive and tries to facilitate a particular behavior without enforcing it. It focuses on informational instruments such as endorsing statements, guidelines, or websites and utilizes non-obligatory economic instruments like subsidies or labels [105].

6.2.2 Self-Regulation by Businesses

When businesses specify rules for themselves, they usually do so to ease or pre-empt other forms of regulation. Therefore, self-regulation is usually not isolated from but driven by (discussions about) other forms of regulation. For example, when the US Chemical Manufacturers Association noticed a decline in public opinion about the chemical industry,

they launched the Responsible Care program to appear more trustworthy. This decision almost certainly aimed to prevent a more binding form of regulation by the government or civil society.
Self-regulation occurs in two forms: Collective regulation refers to a group of significant companies establishing agreements, standards, or codes of conduct applied to the industry. Companies are usually not forced to participate in the regulation initiative. Still, non-compliance could lead to sanctions such as being excluded from the initiative, leading to economic disadvantages [105]. An example in the AI industry is the "Partnership on AI", which counts 95 partners such as Amazon, Apple, Google, Meta, Microsoft and others [106].

Individual regulation means that a company is binding itself to some form of regulation by establishing management systems, Corporate Social Responsibility (CSR) strategies, or codes of conduct. Such individual self-regulation can often be a response to business partners' demands for certain CSR practices. Nike poses an example as it launched a supply chain audit regime on labor conditions to respond to public pressure, backed by large suppliers, bulk buyers, and institutional investors [105].

6.2.3 Regulation by Civil Society

As the example of Nike shows, public opinion can have an impact on businesses' decisions. Civil Society Organizations (CSOs) can leverage this by setting formal standards or putting pressure on companies.
With formal standard-setting, the CSOs develop standards that help steer businesses in the desired direction. Usually, these standards target globally operating companies and, therefore, are international. An example of such formal standard-setting is found in the label "Rugmark" which a CSO defined to certify carpets made without child labor [105].
Opposite the formal approach, CSOs can also invoke regulation by informal pressuring. For this, CSOs mobilize stakeholders to confront businesses with social or environmental claims. These claims are often based on moral claims or legitimacy in the eyes of the public. Such confrontation by stakeholders can motivate companies to change their practices to protect their resources & productivity [105]. An example of this happened in the AI industry when Google withdrew from the military project "Maven" after protests from its employees. The project aimed to analyze drone footage using Artificial Intelligence, which would ultimately result in the killing of humans using unmanned vehicles [107]. A similar case happened with Microsoft when over 300 employees protested against the company's cooperation with Immigration and Customs Enforcement (ICE) [108]. This shows that civil society can effectively help regulate certain cases, but the effectiveness of the regulation depends on how many essential stakeholders the CSOs can mobilize[105].

6.2.4 Co-Regulation by a Collaboration of Actors

While each actor alone has some options to invoke regulation, collaborations among two or all actors expand the typology of regulation. Collaboration in the private sector – **private co-regulation** – includes voluntary certifications defined by businesses and civil society organizations, such as the Forest Stewardship Council or the Marine Stewardship Council. Co-Regulation between companies and the government is called **public co-regulation** and manifests in negotiated agreements. A collaboration between the government and civil society usually focuses more on managing common-pool resources. Because of this, it is called **public co-management**. A collaboration between all three actors – the **tripartite co-regulation** – includes certification schemes and partnerships. The actors strive for common goals and use synergies by sharing domain-specific resources, almost like in a network. An example provides the Tripartite Climate Change Partnerships in the UK [105].

6.3 When to regulate? - The Challenge for Technology Regulation

To efficiently regulate something, one must not only know how to put regulation in place but also when to start regulating - especially because rule making is time-consuming. Regulators need to gather the relevant and correct information about the entity to be regulated and decide on the form of regulation – all while not regulating too soon and not too late. For this, they rely primarily on facts that experts present. The problem with such fact-based approaches is that they reach their limits when innovation cycles are quicker than the time needed for gathering facts. The pressure of time in those cases means that there might be not enough or "wrong" facts about the innovation to be regulated to make an educated decision about when and how to regulate. Yet, simply waiting to or being overly cautious in creating a regulatory framework can have a heavy economic price. "First mover" markets will profit from benefits that offering an innovative service or technology provides. Taking too long to regulate a particular technology can also result in the regulation already being outdated once it comes into effect. For example, rule-makers could restrict using people's faces to train algorithms, but algorithms are already being trained by artificial facial data [109].

The long regulation cycle and the time-pressure on gathering facts to decide when to implement regulatory means imply that facts need to be gathered as early as possible and from objective sources. An approach to how this could be achieved is by gathering investment data. Fenwick et al. suggest that the point in time where early-stage investments peak and later-stage investments start growing is the right time to begin implementing regulatory means. The investments mean there is an interest and value to the technology

and that it almost certainly will make it to market [109]. Figure 15 provides an insight into investments in AI and implies that regulation on AI technology should have been taken up between 2011 and 2012.

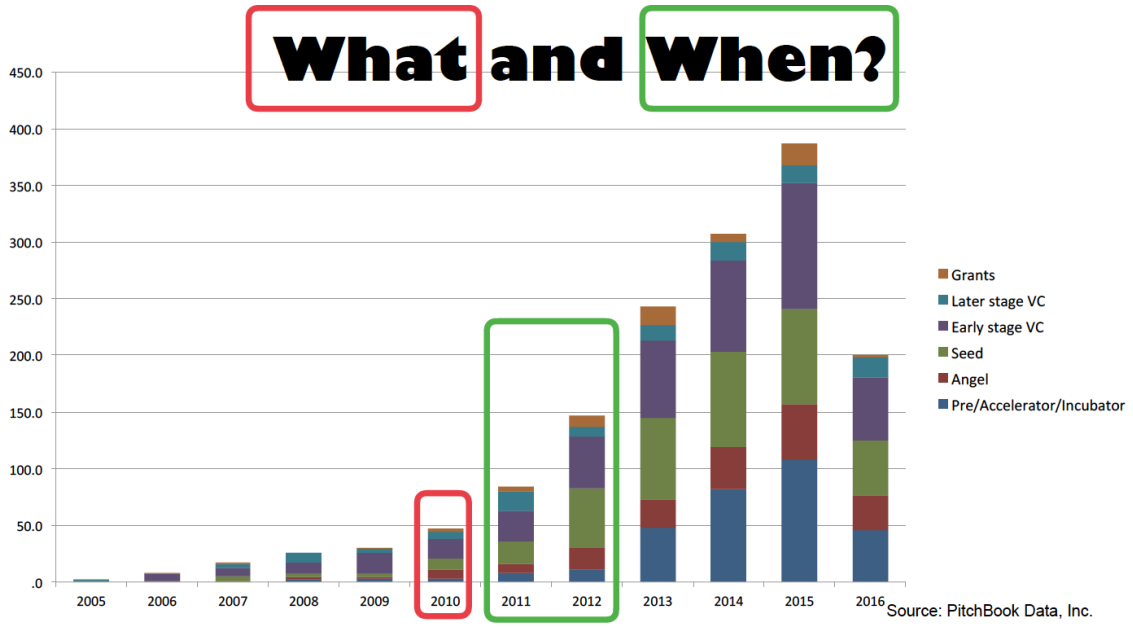


Figure 15: Venture Capital Investments in Artificial Intelligence showing when the right time would be to start regulation debates [109].

Another problem of contemporary regulation procedures is that regulators often think of regulatory practices as one-time matters. But with rapidly and continuously evolving technologies like Artificial Intelligence, there is a need to rethink how often regulation should be discussed. Fenwick et al. present a principle-based approach to enable revisions of the regulatory framework to react to new knowledge or advancements of the technology [109].

This section has shown that there are three actors that can enact regulation. Regulation mostly uses instruments that create economic incentives like taxes, subsidies, labels or certification schemes. Only legal instruments do not directly impact the economic success of regulated companies. Even civil pressure will result in an economic incentive when the reputational loss of the company has a greater financial impact than the economic success of the malpractice. Furthermore, the section has displayed that regular regulation practices are not suited for regulating new technology. The traditional process of defining regulation is slower than the innovation cycles of new technologies. Because of this, regulation is at risk of being outdated once it comes into effect. Therefore, the next section will look at how governments are approaching regulation of AI.

7 Current Situation of AI Regulation

Countries all over the world have begun to implement approaches that aim to regulate the applications of Artificial Intelligence in different ways. This chapter will describe what kinds of regulations have been implemented and look at what the government of Switzerland is doing in this field.

7.1 Regulatory Situation of AI in leading Countries

The Future of Life Institute and the OECD.AI Policy Observatory provide an overview of international and national AI strategies. The Future of Life Institute lists six international agreements and initiatives from the United Nations, European Union and the G7 leaders, among others. Most of these international initiatives consist of an agreement between the participating countries to collaborate on the development of AI and its guidelines.

In terms of national strategies, the Fire of Life institute has identified national AI strategies for 36 countries [110]. The OECD.AI Policy Observatory lists 52 countries with national AI policy initiatives, excluding Malaysia and Tunesia which are present in the Future of Life Institute's list [111]. Figure 16 provides an overview of the countries on a world map. The darker a country is displayed, the more initiatives it has launched.

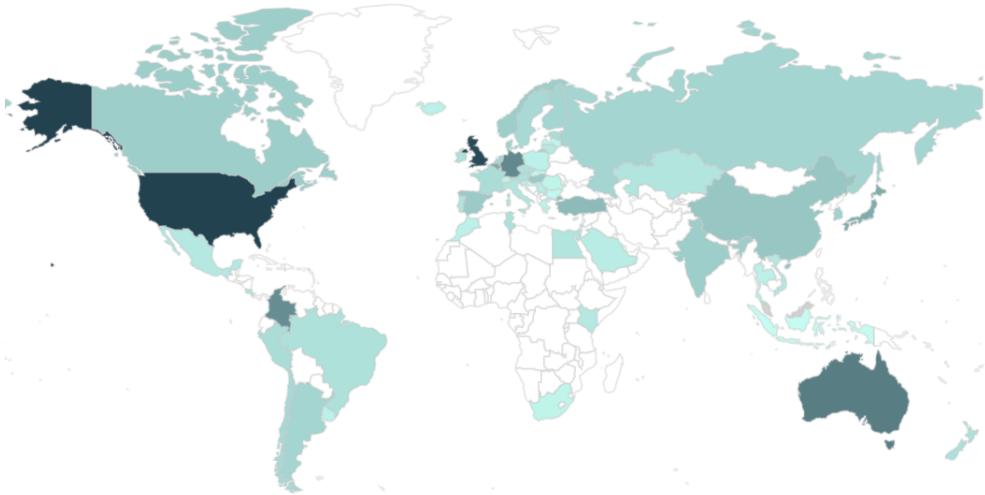


Figure 16: National AI initiatives around the world [112]

It's evident that many countries have realized the need to participate in the development of AI and to reap its benefits. As is normal with emerging technologies, some countries are further along in adopting the technology than others. The Oxford Insights' Government AI Readiness Index ranks governments across the world according to their readiness to implement AI. It's important to note that this refers to the implementation of AI in the

delivery of public services to the citizens. Table 4 provides an overview of the Top 20 countries according to the Government Readiness Index.

Rank	Country	Score	Rank	Country	Score
1	United States of America	85.479	11	France	73.767
2	United Kingdom	81.124	12	Australia	73.577
3	Finland	79.238	13	Japan	73.303
4	Germany	78.974	14	Canada	73.158
5	Sweden	78.772	15	Luxembourg	72.616
6	Singapore	78.704	16	United Arab Emirates	72.395
7	Republic of Korea	77.695	17	Estonia	69.922
8	Denmark	75.618	18	Switzerland	69.219
9	Netherlands	75.297	19	China	69.080
10	Norway	74.430	20	Israel	68.825

Table 4: Top 20 AI-ready countries [113]

In 2020, the AI-Readiness Report did not only include a general index but also included a ranking of the "responsible AI sub-index". For now it evaluates 34 governments, comprising on the Top 20 as seen in table 4 along with regional leaders mentioned in the report, on how responsibly they use AI. It measures nine indicators across four dimensions: Inclusivity, Accountability, Transparency and Privacy. It is important to note that these indicators do not take into account the specific regulations proposed by the respective nations in their AI strategies but rather relate to properties of the governments such as tendency to corruption and surveillance, power of lobbying or the ability to hold the government accountable [113]. Table 5 displays the ranking for the Sub-Index.

Rank	Country	Score	Rank	Country	Score
1	Estonia	79.852	11	Canada	65.005
2	Norway	77.201	12	Netherlands	64.472
3	Luxembourg	76.526	13	Mauritius	64.099
4	Finland	76.172	14	Switzerland	63.653
5	Sweden	72.975	15	Luxembourg	60.872
6	Portugal	72.436	16	Germany	59.723
7	New Zealand	68.262	17	Romania	58.997
8	Denmark	66.873	18	Australia	58.771
9	Senegal	66.381	19	Singapore	57.548
10	Uruguay	65.205	20	France	56.666

Table 5: Top 20 of the Responsible AI Sub-Index ranking [113]

It's interesting to see that the two most "AI-ready" governments are not featured in the Top20 of the Responsible AI Sub-Index. The United Kingdom is ranked as number 22 and the United States of America is found on rank 24. However, this should not lead to

the hypothesis that the indices are correlated, because there are countries that rank high in both indices, such as Finland, Sweden and Norway.

In order to provide structure to this section, the following paragraphs will provide an overview of policy initiatives taken by the top 5 countries in each ranking along with, if applicable, the regulatory aspects covered in their respective AI strategies.

Estonia. Estonia, along with Denmark, Finland, Faroe Islands, Iceland, Latvia, Lithuania, Norway, Sweden and Åland Islands signed a declaration to, among other goals, collaborate on "developing ethical and transparent guidelines, standards, principles and values". It has also launched several initiatives on open data models and open AI components alongside its national AI strategy. This strategy sees some changes necessary in different laws and announces a legislation bill for enabling the uptake of AI [111] [114].

Norway. Norway also sees an open data approach by launching a national data catalog and it endorses the use of sandboxes to test and nurture AI applications in a suitable environment[111]. Its national AI strategy foresees that the government will "review and assess regulations that hamper appropriate and desired use of Artificial Intelligence in the public and private sectors" and that it will set requirements for transparency and accountability for AI systems used in public administration[115, p.28].

Luxembourg. Luxembourg has passed a strategic vision document for AI, which sees it becoming "one of the most advanced digital societies in the world" [116, p.5]. In terms of regulation and ethics, the government wants to invest in an AI-friendly framework and remove barriers to secure AI development. The strategy also ensures the implementation of legal and ethical guidelines to protect fundamental rights. No other national strategy paper explains the will to include the public's opinion into its policymaking as clearly as Luxembourg, which indicates that the public's trust in AI technologies will play a role in how restrictive AI policies will be [116, p.24].

Finland. Finland has passed a national regulation on automated decision making, which unfortunately is not available in English. Together with Amsterdam, Helsinki has also launched an AI Register, which provides a way to track how algorithms are being used in municipalities [111].

Sweden. Sweden's national approach to Artificial Intelligence contains a section on framework and infrastructure which describes the need for Sweden's government to "develop rules, standards, norms and ethical principles to guide ethical and sustainable AI and the use of AI" [117, p.10]. It also acknowledges the importance of European and international frameworks, such as the GDPR which it describes as an important part of the AI framework. No other government has described the need for early regulation through clear guidelines and standards to guide development in a beneficiary direction as clearly as Sweden has [117]. Sweden also has put in effect a committee for technological innovation

and ethics to help identify policy challenges and accelerate policy development not only in AI technology but all technologies relating to digital transformation [111].

United States of America. The AI strategy from the USA differs in tone from the AI strategies of other governments. It puts little focus on regulating AI for ethical reasons and focuses more on what needs to be done in the security department to keep up with other countries' (specifically China's or Russia's) advancements and protect the nation's technology from being "stolen" by these countries[118]. Outside of the AI strategy, the USA has launched 50 initiatives, including initiatives on automated vehicles and NIST Principles for explainable AI [111]. One of those initiatives is the Algorithmic Accountability Act. It was initially drafted in 2019, revised based on discussions with experts in the field and re-introduced to the senate in February 2022. The Act focuses mainly on creating more transparency about when and how algorithms are used and requires companies to assess impacts of their automated systems [119].

United Kingdom. The United Kingdom has defined governance and regulation as one of three pillars constructing its AI strategy. The strategy announces the release of a white paper in 2022 which will define the UK's position on governing and regulating AI. Depending on the results of that white paper, the strategy introduces three alternative options to policymaking: (1) Remove existing regulatory burdens if they create unnecessary barriers, (2) retain the existing sector-based approach, ensuring the regulators are empowered to make decisions or (3) introduce additional cross-sector AI-specific principles or rules to support individual regulators. Until then, it stresses the focus on a pro-innovative approach and promises to engage in international policymaking [120]. Along with the strategy document, it has also released a regulatory framework for automated vehicles, a review into bias in algorithmic decision-making and a data ethics and AI guidance landscape, among others. It's important to note that Scotland and Wales have also released their own AI strategy, alongside the UK strategy [111].

Germany. Germany's strategy recognizes that standardization and norms can help build the public's trust in AI technology. It describes the intention to evaluate whether current laws, especially in product liability and product safety, build a sufficient legal framework for AI systems. The document also mentions the importance of legal certainty and the plan to implement incentives for innovations in safe and trustworthy AI. It also briefly discusses bias and discrimination in algorithms and proposes a solution by building socially diverse development teams [121].

China. Although China is ranking last according to the Responsible AI Sub-Index, this thesis will briefly discuss their AI Plan, which was released in 2017, as China is seen as

an important global player by other countries and international organizations, especially the USA, whose AI report mentioned China's ambitions several times [118]. China's "Next Generation Artificial Intelligence Plan" sees it becoming an AI world leader by 2030. While it clearly describes the areas in which it wants to apply AI, for example Smart Courts, it follows the example of other countries' strategies by also stating goals that relate to AI regulation. It describes the desire to establish norms and AI standards in order to provide the grounds for "a healthy and rapid development of AI". Few other governments have defined their plans to regulate AI as clearly as China has. It intends to establish laws and mentions areas in which such laws will be needed, for example autonomous driving or service robots. On March 1st, 2022, China has put the Internet Information Service Algorithmic Recommendation Management Provisions into effect. This regulation outlaws filtering and recommending content, dynamic pricing or sorting search results through recommendation systems that base their decisions on personal data. Companies are also required to inform users when recommendation systems are used and provide a method for users to opt-out of such systems. While China seems to move very quickly in terms of AI regulation, government applications are not subject to this regulation[122]. Like other governments, China also states that it wants to participate and play an active role in the shaping of international discussions about AI [123].

European Union. The European Union is the first and so far only international organization that has released a proposal for an International AI Act. The proposal is a response to several member governments releasing their own AI strategy and intentions for regulations, as described above. The EU sees a problem in these separate national AI rules as they can lead to fragmentation of international markets and substantial diminishment of legal certainty for providers and users of AI systems. It is the only regulatory framework that foresees the prohibition of AI systems that have the potential to affect fundamental rights or to manipulate persons. Additionally, it requires so-called "high-risk AI systems" to comply with certain mandatory requirements and undergo conformity assessments before being permitted on the European market [124].

AI has become a globally relevant topic, as evidenced by the many AI national strategies that have launched in the past years. Most of these strategies mention regulatory intentions but the means through which these intentions should be fulfilled differ. Some countries like Sweden, Estonia, the EU and China propose to create new or adapt existing laws according to AI's challenges. The governing bodies agree that regulation should not hinder research and development of beneficial AI applications but few actually propose a way to distinguish profitable from harmful AI applications except for the EU, which categorizes applications based on risk, and Norway which proposes a sandbox environment to test AI applications, so harmful consequences won't be as devastating. The degree to which the public will

be able to influence regulation is mentioned in few strategies, but the majority does not mention this aspect. A common goal that is mentioned in all strategies is the will and desire to participate in international discussions about regulations and shaping the future of AI.

Since most of these strategies are very young, their fulfillment or success cannot be analyzed at this point in time. It remains to be seen if the intentions set by the governments are actually realized.

Given that many countries have defined their AI strategies and put in place certain initiatives on AI, it is important to have a look at how Switzerland plans to regulate AI in the future.

7.2 Current Regulatory Situation of AI in Switzerland

Having looked at the different approaches from governments around the world, it is now of interest to see how Switzerland approaches the challenges posed by AI and how it compares to other countries in this regard.

Reactive regulation. The Swiss Government also seems to have recognized the need for a legal & regulatory framework that prevents possible threats coming from AI. This is expressed through the creation of an AI task force, which released a report in December 2019, identifying the challenges AI poses for Switzerland. Additionally, the report evaluates whether existing laws would be able to handle the specific challenges. The report claims that the Swiss legal framework is sufficient for the current state of AI and no adaptions need to be made. However, the task force stresses that the situation of the legal framework regarding AI needs to be continuously evaluated and actions need to be taken should applications arise that remain unregulated [125].

AI Strategy. Unlike other countries (as described above), Switzerland has not drafted or passed a dedicated AI strategy. However, in September 2020, the Swiss government has released a Digital Strategy paper that describes the goals related to digital transformation that the country wishes to achieve. This strategy includes an action field on "Data, digital Content and Artificial Intelligence", the objectives of which largely relate to the secure generation and use of (personal) data. The only AI-specific goal reads "The general conditions for the transparent and responsible application of Artificial Intelligence are optimized"[126]. Since the description of this goal bases on the previously described report on AI challenges by the government task force, it's not surprising that it also sees a need for the constant monitoring and evaluation of undesired consequences by algorithms. It mentions that it's necessary for Switzerland to take international, and especially European, developments into account when drafting new rules for AI. This implies that the Swiss

government is generally open to defining rules related to AI applications. According to the strategy, especially algorithmic decision systems need to be transparent and verifiable and respect active laws[126]. Overall the strategy endorses digital transformation and describes efforts to support businesses in becoming more digitized or even launch new digital business models. It describes the will to provide favorable conditions for businesses with a digital focus to thrive in Switzerland and to stay an attractive location for innovative players. As was prominent in the report on AI challenges, the strategy also avoids mentioning or promising support for certain technologies and follows the same technology-neutral approach[126].

Competence Network for Artificial Intelligence. On August 25, 2021, the Swiss Federal Council agreed to develop a competence network for AI. In spring 2022, the Competence Network for Artificial Intelligence (CNAI) was launched[127]. Its goal is to foster the use of Artificial Intelligence within and outside the Federal Administration. It consists of two communities. The community of practice is open to the public and aims to foster the exchange of knowledge and best practices. This community is closely related to the Data Science Competence Center's "data science for good" community. The community of expertise consists of experts curated by the CNAI. The members are working for the Federal Administration, have technical expertise, and can be consulted on concrete project issues. Besides fostering the use of AI, the CNAI also pledges to contribute to informing the public about the technology. One step to do this is to ensure the transparency of AI projects through a public project catalog[128]. This catalog lists 25 projects in three Federal Departments[7][129].

Digital Trust Label. In January 2022, the NPO "Swiss Digital Initiative" has launched he Digital Trust Label[130]. The goal of the label is for users to better understand what happens with their data when they are using digital services and that they feel safe and secure when using these services. It encompasses 35 criteria across four dimensions: Security, Data Protection, Reliability and Fair User Interaction. The most AI-specific criteria are found in the latter dimension. Examples include the following[131]:

- "Are service interfaces designed so as not to deceive or clearly manipulate users?"
- "Does the user understand when AI-based algorithms - especially decision-making algorithms - are used?"
- "Does the service provider indicate which user-related data is processed by Artificial Intelligence, and why?"

[7]Federal Department of Foreign Affairs FDFA, Federal Deparment of Home Affairs FDHA and Federal Department of Defense, Civil Protection and Sport DDPS

Technology-neutral approach. Switzerland praises itself on its technology-neutral approach to phrase laws and policies. This neutrality should ensure that companies and research facilities are not deciding to use or research a certain technology under government influence [125].

AI in the public sector. Based on the above-mentioned report by the task force, the Swiss government has released guidelines on how AI should be used by federal agencies and external partners with governmental tasks. The guidelines should serve as a general frame of reference and ensure a coherent policy. Therefore, these guidelines are no regulatory framework in themselves, but it is specifically mentioned that they should be considered when adopting specific regulations and when helping to shape international regulatory frameworks. Against this background, it is important to know what the guidelines encompass. The guidelines consist of seven values:

1. Putting people first
2. Regulatory conditions for the development and application of AI
3. Transparency, traceability and explainability
4. Accountability
5. Safety
6. Actively shape AI governance
7. Involve all relevant national and international stakeholders

Guideline 1 focuses mainly on protecting human and fundamental rights, such as self-determination and equality. It encompasses the intention to implement safeguards and controls for AI applications that could potentially affect fundamental rights. In guideline 2, the government manifests that the technology-neutral approach for regulation should continue so that business and science stakeholders may decide independently and uninfluenced which technology to apply and promote. They also manifest in this guideline that "regulatory conditions for digital solutions should be designed in such a way that does not hinder innovations and contributes both to greater value creation and sustainable development" [132, p.4]. Like many other countries, Switzerland sees the values stated in the third guideline as prerequisites for trustworthy AI. An important part of the guideline is that people need to clearly recognize if they are interacting with an AI or if decisions that affect them are being aided or made by AI. These decisions also need to be comprehensible to the affected person. Another key aspect of this guideline is the disclosure of the AI's code and training data. The guidelines of accountability and safety require having

a clear definition on who holds liability for the AI (since liability must not be delegated to machines) and that AI must not be susceptible to abuse. Finally, the guidelines manifest that Switzerland should actively work with other nations to shape global AI governance and that the opinions of national and international actors and target groups are considered when making decisions about AI. As a final goal, the document plans to establish Geneva as a center for digital governance of AI [132].

Because of the Swiss federalism principle, cantons can also launch their own AI initiatives. In March 2022 the canton of Zurich announced an Innovation-Sandbox that will serve as a testing environment for AI use cases. It allows to implement selected projects in a controlled framework, for example algorithms that predict traffic jams or diagnose rare diseases. The sandbox benefits participants through offering support for regulatory questions and access to public data sources. Simultaneously, the government of Zurich can build competencies and insights into how to regulate AI. The sandbox will launch the first projects in summer 2022[133].

AI in the private sector. Since the government has not yet taken an active role in regulating AI in the private sector, Swiss companies currently are self-regulating. A report by the Swiss Academy of Technological Sciences SATW stated that stability in regulation is important, especially for start-ups who need legal security. This might be a problem since the current approach by the Swiss government seems to be a more reactive approach [134].

Proposed actions. While criticism of the Swiss Government's AI governance is scarce, a few proposed actions from other parties than the task force can be identified.

Although the task force concluded that the legal and regulatory framework is sufficient to address AI's general challenges, it has asked government agencies to propose recommendations for actions in their respective AI application areas. Notable proposed actions are as follows [125]:

- Creating a legal framework for automated mobility
- Evaluate new kinds of threats, i.e. cyber security, propaganda, warfare
- Follow up on the development of premiums using AI in state-certified private insurers
- Develop data protection laws

These proposals should be evaluated carefully since it is not made fully clear in the report who was involved in making the proposals and personal interests, lack of expertise or fear of new technology might have played a role in what should be proposed and what not.

It is interesting, however, that the department of mobility has proposed the creation of a special legal framework for automated mobility when the task force did not see a need for adapting any laws.

It also needs to be noted that the task force has identified two important action fields of AI that need consistent monitoring: Switzerland should participate in international efforts to create AI-specific rules through Industry standards and ethical AI principles. Also, the use of AI in media should be carefully monitored since AI applications for commercial and political goals can influence public opinion.

In general, the Swiss government's approach to regulating AI corresponds to the approaches of other governments described above. It also has acknowledged the potential and challenges of AI and has put in place certain guidelines for AI. However, there are no consequences described if somebody should violate these guidelines. Switzerland's strategy agrees with other governments to define regulatory measures in such a way that they do not hinder beneficial research and the establishment of new businesses. Like all other governments covered in this thesis, Switzerland also plans to participate in international discussions. Where the Swiss approach differs from other governments' approach is in the more passive approach. As other governments already state the need to adapt or establish laws and to educate policymakers on AI, Switzerland does not mention any actions of this sort. It also seems as though Switzerland does not put as much effort into AI as it does in Digitization in general. Other governments have stated plans and established visions to become continental or global leaders in AI technology, but the Swiss strategy lacks any vision of this kind. Since AI applications in Switzerland are mostly developed in the health industry, which is known for its sensitive data privacy requirements, [125, p.46] it's important for Switzerland to have a stable regulatory framework to advance the field of health care in Switzerland by developing AI applications for the industry.

8 Learnings from Related Fields

Every time a new disruptive technology emerges, initial excitement eventually becomes clouded with questions about ethical challenges and discussions about regulatory approaches. This section will look at the regulations and regulation attempts relating to three disruptive technologies. With the exception of biological weapons, the technologies covered were selected mainly because they promise major benefits to society but also pose challenges that could result in harmful activities.

Biological Weapons. The regulation of biological weapons started in 1925 when countries agreed to refrain from using biological weapons [135]. Because this agreement did not effectively prevent attacks, the Biological Weapons Convention (BWC) was held in 1972. It serves as the main regulatory milestone of biological weapons and is defined as a treaty that prohibits the use, development, production and stockpiling of any biological weapon in unjustifiable quantities. Researchers are criticizing the BWC for failing to provide inspection and verification processes, therefore allowing signatory parties to continue acting unsupervised. Since the BWC did not provide any guidelines on enforcement or how to deal with violations, attacks using biological weapons were still carried out by participating nations [136][135]. Under the BWC, research on biological weapons remained allowed. This proved useful as dangerous biological agents could be transformed into beneficial drugs. Anthrax for example was found to be able to help cure cancer and the deadly agent code-named "Agent X" is now used as Botox [137][138].
This example indicates that research in a controversial field can transform dangerous agents into beneficial applications if the right intent is driving the research. However, the BWC's approach to regulation through an international agreement indicates that agreements to guidelines without a clear plan of enforcement and control are ineffective as many countries have violated the agreements without consequences [136].

Human genome editing / CRISPR. As with all emerging technologies that invade the human body, the advent of human genome editing and synthetic biology raised public concern. Although the research community recognized that alongside striking benefits, the technology also posed ethical challenges and possible negative impacts, the existing regulatory framework for biotechnology product development was considered sufficient. Despite the consensus being that "regulatory options should support innovation and commercial development of new products while protecting the public from potential harms" and certain nations having released guidelines adhering to ethical challenges, these measures are being violated without consequences [139]. The possibly most prominent case of violation is one of He Jiankui, a scientist who used genome editing on babies to make them immune to the HIV virus. Reactions of the science community to this violation express the imperfections of such self-regulatory, uncontrolled and agreement-based approaches.

Members of the community stated that they "had no authority to stop him" and described the incident as a "failure of self-regulation" [140]. In light of this event, some researchers called for an international oversight group, proposing the UN as an acting body. The community remains divided over whether such an oversight body would have been able to stop the experiment, as He had already talked to several researchers before and during his experiment, and none were able to stop him [140]. Three months after the conference where He presented his case, 18 researchers, including Emmanuelle Carpenter who won the Nobel Prize for her CRISPR, called for a temporary moratorium on human genome editing, so there would be time to hold discussions about technical, scientific and ethical issues, as well as time to establish a regulatory framework [141].
This example enforces the previous argument that international agreements without enforceable measures in the event of violation do not serve as effective regulatory measures. In contrast to this experiment, a few decades earlier, a scientist had conducted pioneering work in gene therapy. He did so outside his home country, as he believed approval for his project would take too long. The scientist explains that he believed his experiments set the field back several years as his experiment prompted many regulations and oversights in the field of gene therapy [140].
This example indicates two important aspects of regulation: First, that national policies can cause researchers to choose where they practice based on regulatory conditions, which can lead to international tensions. Secondly, it points out that problems can arise when research moves faster than regulation, and that regulatory actions need to keep pace and may need to be proactive to prevent development from going in undesirable directions and to protect confidence in the technology.

Distributed Ledger Technology / Cryptocurrency. Distributed Ledger Technology (DLT) is arguably the closest related technology to Artificial Intelligence in this sample as it's also located in the field of computer science. It mostly differs from the previously explained technologies in terms that blockchain regulation does not rely on international agreements but rather on existing laws, for example money laundering laws.
Cryptocurrency can provide massive benefits but also be abused for tax evasion, money laundering or terrorism financing. Shanaev et al. suggest that excessive regulation can harm cryptocurrency as it drives prices down but endorses governments to "let the industry develop in a freer environment", so market volatility is lowered and prices are more stable [142]. Such a free environment could be described as a "sandbox" approach, which is implemented by the Swiss government. Binder et al. support the Swiss government's approach and identify it as a reason for Switzerland having established itself as a prime location for DLT-applications [16]. Unfortunately, Switzerland is doing little to minimize criminal activity using cryptocurrency. A Swiss government report finds that current money laundering laws are technology-neutral enough to cover crimes involving cryp-

tocurrencies [143].
While governments seem to rest upon the existing laws to enforce legal action against criminals, it must be considered that these laws were most possibly written under the assumption that sender and receiver of a transaction are known (through data known to the bank) or at least traceable (through numbered bills). With DLT however, neither sender nor receiver are traceable and therefore the law might apply but cannot be enforced.
The learnings that can be drawn from DLT regulation are that existing laws might seem sufficient to regulate challenges posed by new technology. Yet it has to be taken into account that in the case of DLT, existing laws are attempting to regulate existing crimes being committed by using new technology. Therefore, these existing laws do not account for *new* challenges posed by this technology, for example anonymity. Hence it has to be evaluated whether these laws will actually be enforceable in cases of violation and if there is a need to adapt them for specific technology-based challenges.

The analysis of this sample of regulations has shown that there currently seems to be no perfect approach for the regulation of disruptive technologies. The important conclusion is to learn from previous cases and not make the same mistakes by underestimating possibilities, relying on international agreements and self-regulation or assuming current laws will cover all current and future issues posed by the new technology. Table 6 presents an overview of what could be incorporated into an AI regulatory framework (positive aspects) and what should be avoided (negative aspects).

Technology	**Posive Aspects of Regulation**	**Negative Aspects of Regulation**
Biological Weapons	Research & use are regulated differently, and research has more liberty International Agreements can prevent legal uncertainty	International agreements can be and are violated without consequences
Human Genome Editing	Researchers practicing in the field should be listened to when making decisions	Different national regulations can lead to researchers conducting studies in another country Self-regulation is not sufficient in regulating delicate technologies
Distributed Ledger Technology	Sandboxes foster innovative environment, minimize harm, and benefit the countries that implement them	Relying on existing laws to regulate challenges arising from new technologies is insufficient

Table 6: Overview of Positive and Negative Aspects of Regulation in other Technologies

Sandbox approaches have proven to foster an innovative environment for start-ups and benefit the countries that implement them. International agreements are a good starting point to prevent legal uncertainty among multi-nationally acting players and to lay a common ground for policymaking. But without having an overseeing body and the means to perform sanctions, these agreements are effectively nothing more than recommendations. In terms of timing of regulation, a proactive but well-informed approach would be able to prevent research from going in the wrong direction. If this is not possible or could lead to missing out on potential benefits, the quick establishment of a sandbox environment would help explore the technology without risking too many negative effects.

9 Survey Analysis

This chapter describes and interprets the results of the survey, which was filled out by industry players and aimed to gain insights into how corporate decision-makers perceive the topic of AI regulation.

9.1 Sample Description

Between February 22 and March 31 2022, 43 people participated in the survey. Figure 17 displays that close to 90% of these participants are decision-makers with an executive/management position. Two-thirds of them describe their position as having a technology focus, and about one-third have a business focus. Three people identified as technology-team members.

Figure 17: The majority of participants are decision makers.

The sample consists majorly of participants whose company already uses AI in their operations, products, and services. Only eight are not using AI, of which six are planning to use AI in the future (see figure 18).

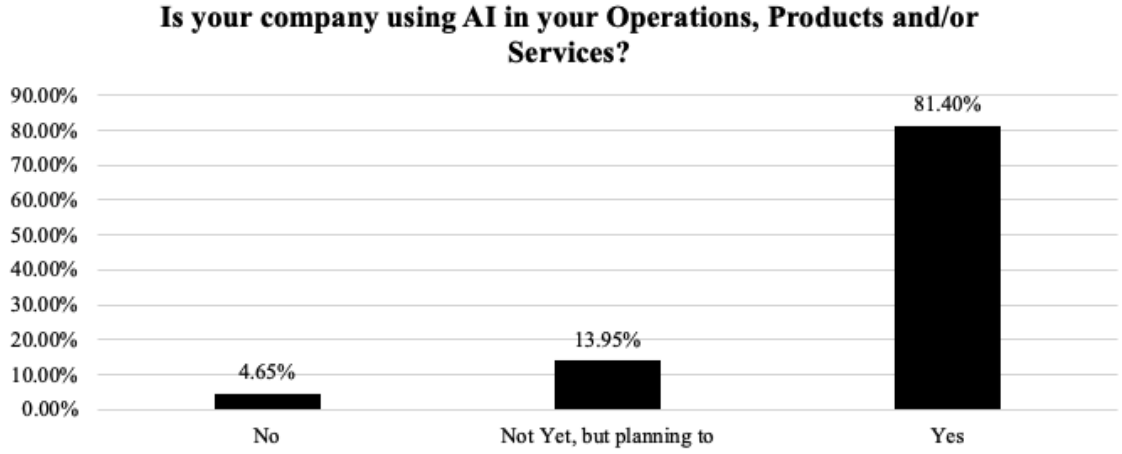


Figure 18: 4 out of 5 participants are already using AI.

As visible in figure 19, almost half the participants work in the IT industry, and about one in six participants works in the Finance industry. Other industries were represented with one to four participants each.
It is essential to consider these statistical properties of the representation when interpreting the survey results.

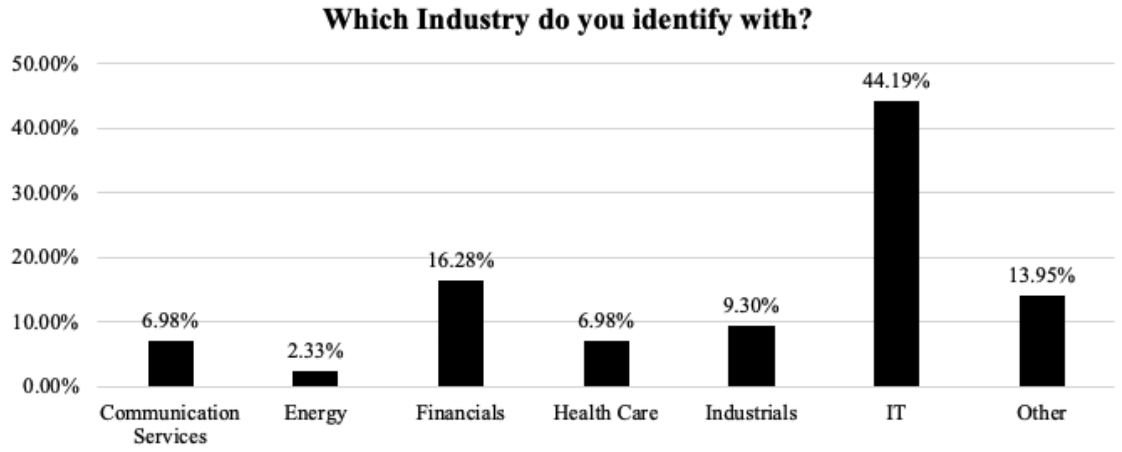


Figure 19: Most participants work in the IT industry.

9.2 To thrive in Switzerland, AI needs more Regulation - The current Situation

The first section of the survey covered the aspect of current AI regulation and asked participants about their perspective on this current situation.

9.2.1 Businesses want more Regulation

In total, 67% of participants favor a more straightforward AI regulation. When asked, "Do we need more specific AI regulation [compared with the current technology-neutral approach]?", only 30% of participants were happy with the current approach. Only one participant wanted less regulation around AI.

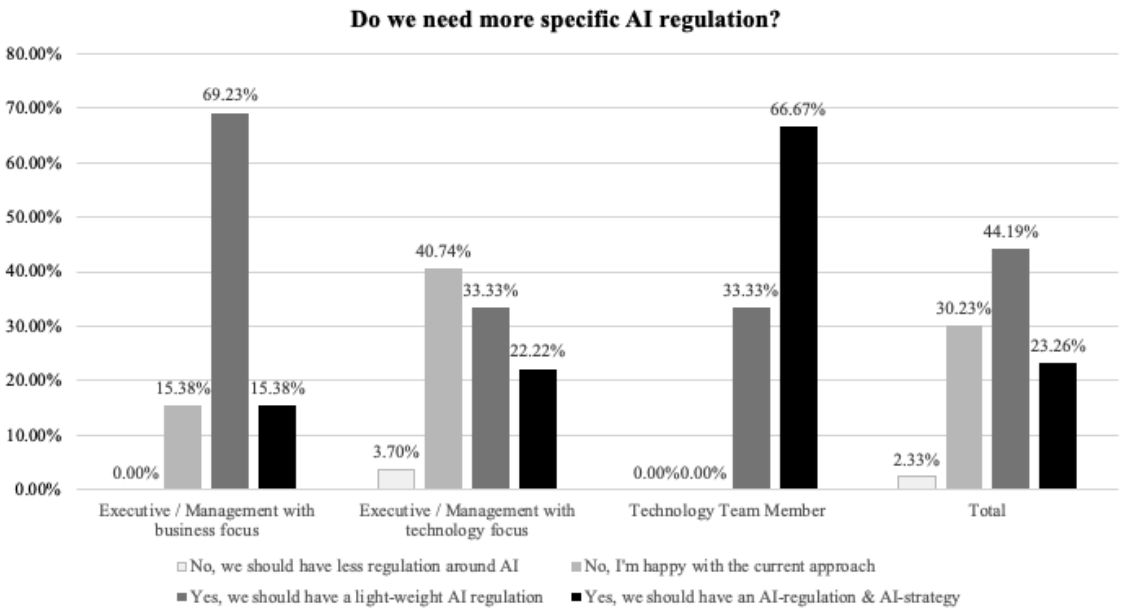


Figure 20: Not many participants are happy with the current regulation approach

Figure 20 shows that an interesting difference emerges when looking at the responses split by position. While 85% of executives with business focus are not happy with the current approach and want more regulation, 40% of executives with technology focus think the current regulation is enough. This difference in perspective could imply that the business executives have a less specific understanding of the technology and would prefer a more "guided approach" to regulation that specifies what needs to be considered when developing or using an AI application. The fact that tech executives are quite divided in what they want could indicate that the company's choice of algorithm and appliance of it influences how the tech executives think about AI regulation.

9.2.2 Self-Regulation is minimal

Only 16% of participants stated that their corporate social responsibility (CSR) policy covered AI aspects (see figure 21). The aspects of AI covered by CSR policies were mainly related to Data Privacy (40%) and ethics (30%). Some individual responses included more practical approaches, such as only using technology that can be explained and transparently reported on or operationalizing AI principles in cooperation with NPOs. About 18% of participants said that they had some other instruments to regulate their use of AI. In total, this adds up to only about one-third of participants having any form of self-regulation implemented in their company.

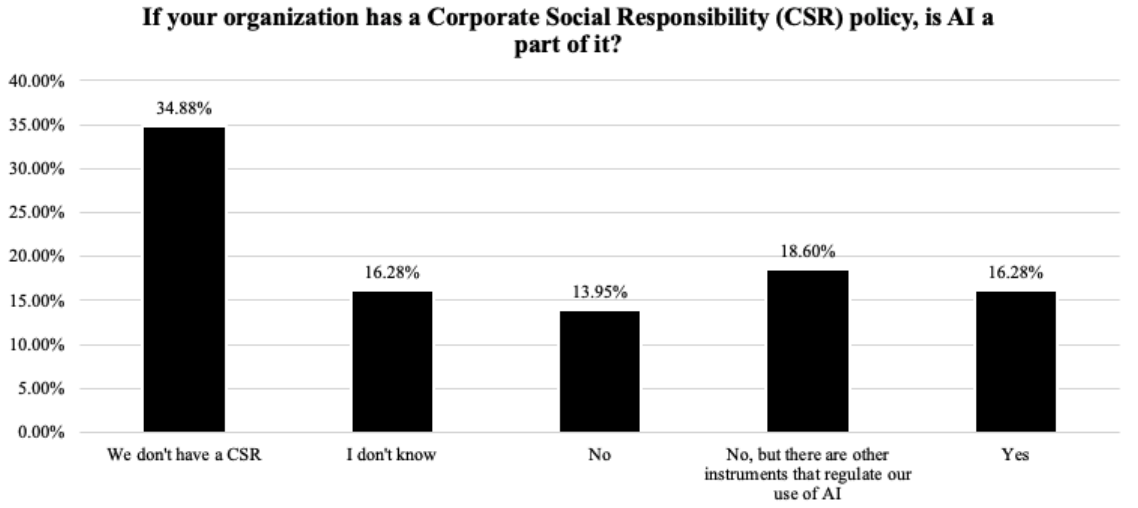


Figure 21: Only one-third of participants' companies have some form of AI self-regulation

Two-thirds of participants either do not have a CSR implemented, or AI is not a part of it. Seven people did not know whether AI was part of their CSR or whether they had any CSR. This could mean two things: Either the CSR is not communicated clearly to everyone in the company or it is not made clear whether AI is a part of it. In any case, if the people in the company do not know the self-regulation measures, they cannot be executed. Combined with the fact that many companies do not have CSR or AI is not covered by it, it can be derived that most companies do not genuinely adopt AI self-regulation.

9.2.3 Ethical & controversial Issues often end in the Project being abandoned

Fifteen of the surveyed companies (37%) had already been confronted with a controversial or ethical issue that arose in developing or implementing an AI solution (see figure 22).

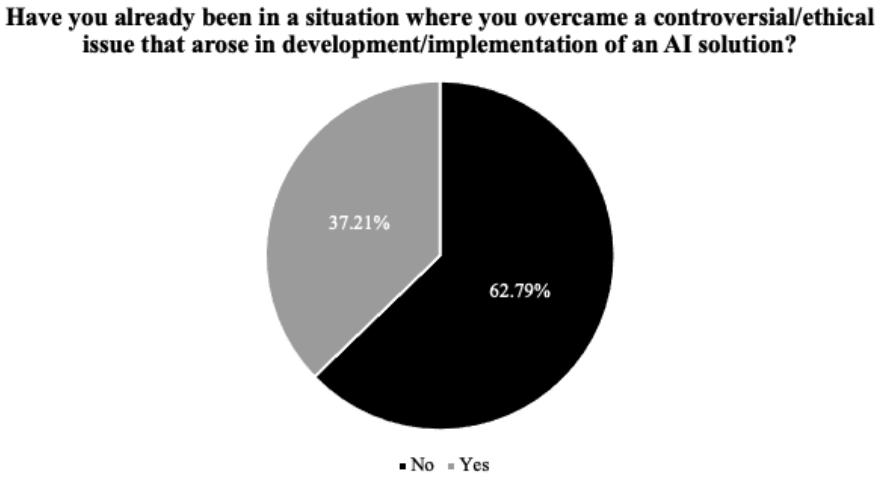


Figure 22: More than one-third of participants already had to overcome an ethical issue

Figure 23 displays that one third of them had to resolve the issue by abandoning the project, removing the feature or not working with the client. Only two companies could resolve the issue by improving the algorithm with more training and testing. This implies that companies are more likely to abandon a project completely than to improve the algorithm when an issue arises. While it is beneficial that ethically unsound projects are abandoned, promising beneficial projects could also end up being abandoned due to issues that could have been resolved through clear regulation. A clear guideline could help companies find a better way to handle such situations or not encounter them in the first place.

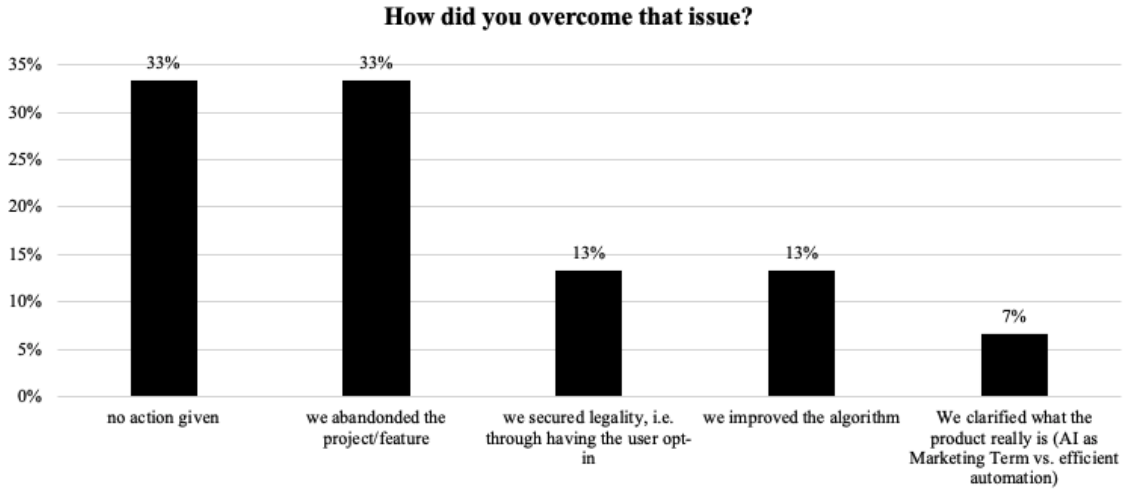


Figure 23: 5 out of 15 companies had to abandon a project when faced with an ethical issue

9.2.4 Most Businesses trust AI Technology

Out of 43 participants, 37 agree or strongly agree that their company should implement more AI in the future (see figure 24). This could mean that businesses recognize AI as a relevant technology and expect it to yield benefits in the future. The fact that 83% of participants who are already using AI endorse having more AI in the future is an indicator of them being happy with what AI is doing for them and that it is proving to be a beneficial and promising technology.

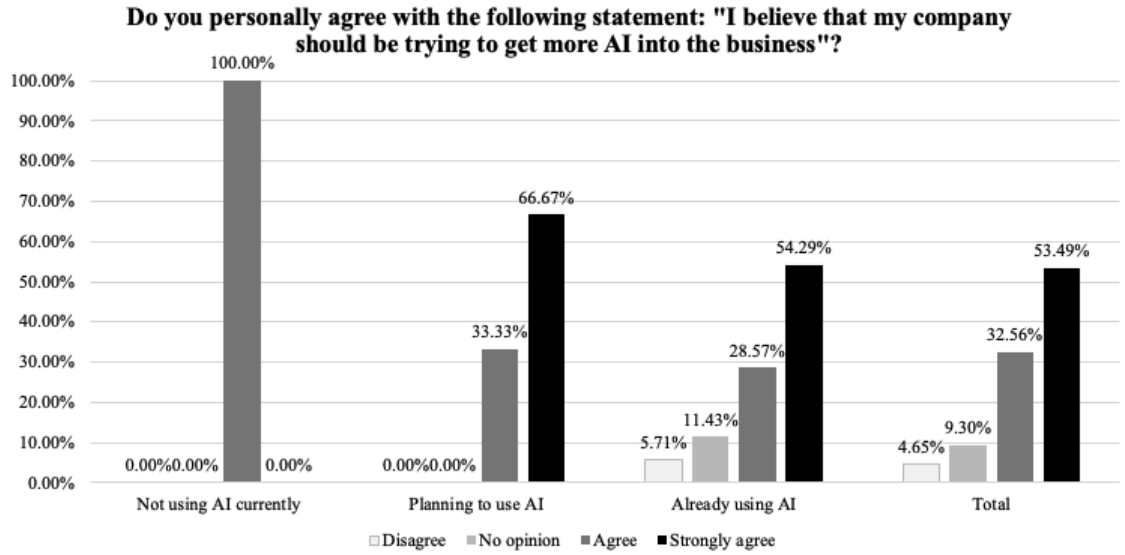


Figure 24: 80% of participants agree or strongly agree that their company should adopt more AI

In this regard, it has to be considered that the survey was mostly sent to companies who are already using AI and could therefore be biased. Companies that are not (yet) using AI are underrepresented and would have to be surveyed more in the future to gain a clearer insight into their outlook.

9.3 There is no Consensus on what AI Regulation should be - The Outlook on Swiss AI Regulation

The second survey section evaluated what the participants think of possible future AI regulation scenarios.

9.3.1 A Label for identifying responsible AI is controversial

The survey participants disagree on whether they favor a label that would identify responsible AI. Figure 25 displays that while 55% would favor such a label, 45% do not.

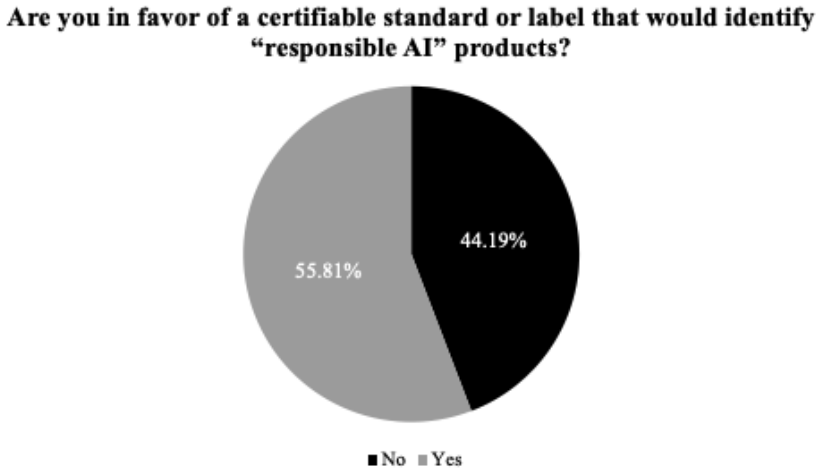


Figure 25: There is no consensus on whether businesses favor a label for responsible AI

Table 7 summarizes the reasons participants mentioned for (not) favoring a label for responsible AI. From these answers, it can be interpreted that businesses do not trust labels that are unclear about their criteria and could be abused as a marketing tool. Additionally, they think that if there is no unanimous understanding of the term "(responsible) AI", the label would already stand on a crumbling foundation or that the possibilities for AI application are too vast for a label to cover all of them.

Reasons for not favoring a label	**Reasons for favoring a label**
We do not have a coherent understanding of (responsible) AI, so a label wouldn't make much sense	Raises & Demonstrates awareness
A label (as well as other forms of regulation) slows down innovation	Demonstrates that available options to act upon ethical issues have been taken
Labels are expensive to maintain	Because AI can lead to discrimination or other forms of harm
Labels are marketing tools or a form to hide behind and do not incentivise responsible behavior	Good for building trust with non-data literate people/end-users
Labels are redundant if we have good regulation	
AI technology is not the problem, its application is - a label couldn't possibly cover all imaginable applications	
Questionable who issues and certifies the label	

Table 7: Reasons for not favoring a label

However, the participants who answered in favor of a label often linked their approval to a condition. These conditions can be summarized as follows:

- Assessments and results need to be made public, otherwise it can be abused for white-washing
- Needs to be "done right" to benefit sustainable development, otherwise it will hinder development
- Needs to be wide-spread and international, otherwise it's useless

Figure 26 shows that, although the total survey sample disagrees on labels, 4 out of 5 companies planning to use AI said they would favor a label. This could imply that a label would provide more guidance for companies looking to build or implement an AI solution.

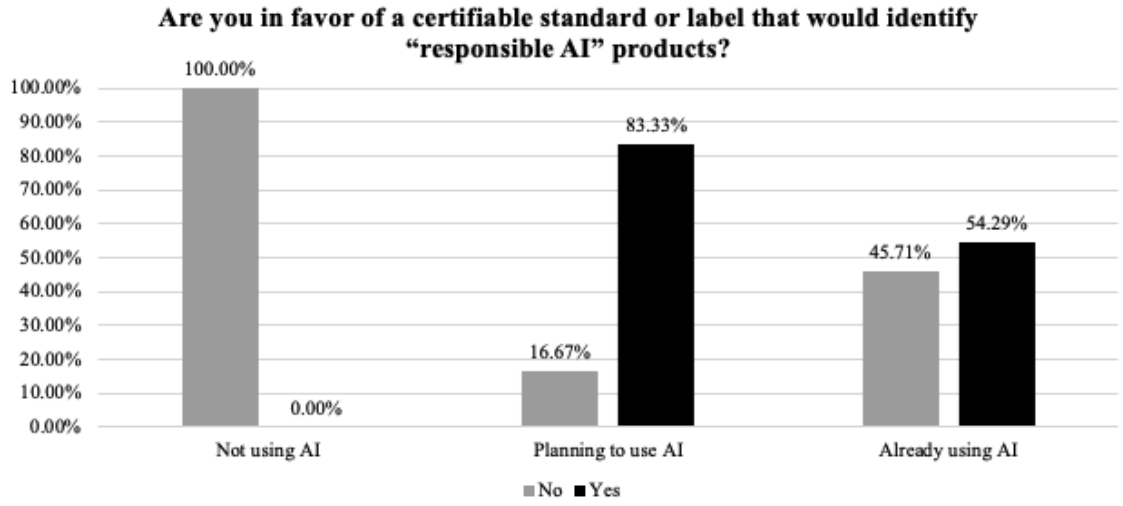


Figure 26: Companies planning to implement AI favor a label more than those already using AI

9.3.2 Business Executives trust more in Self-Regulation than Technology Executives

Half the participants think that AI trust can be built through both governmental and self-regulation. Twenty percent believe that only governmental regulation can strengthen trust and 9% think self-regulation alone would do so. One in five participants thinks that regulation cannot help improve public trust.

Among the sample, executives with business focus believe far more that self-regulation alone can strengthen the public's trust than executives with technology focus (see figure 27). In turn, technology executives are more likely than business executives to not believe in any form of regulation to improve trust. This might be due to the worry that regulation will hinder their autonomy in designing and developing algorithms. They might also believe more in the power of concrete AI applications to foster trust rather than the concept of regulation.

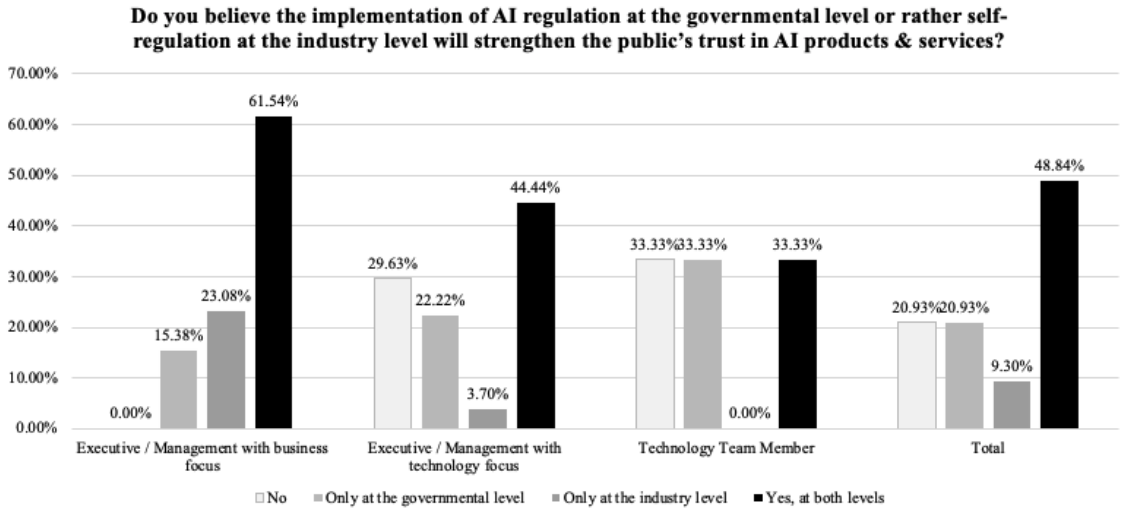


Figure 27: Business Executives believe far more in self-regulation than Technology Executives

9.3.3 A majority of Companies will comply with the EU AI Act

Four in five participants believes that their company will comply with the EU AI Act even if the Swiss approach is less restrictive. This could be due to many companies already being present in or wanting to enter the European market.

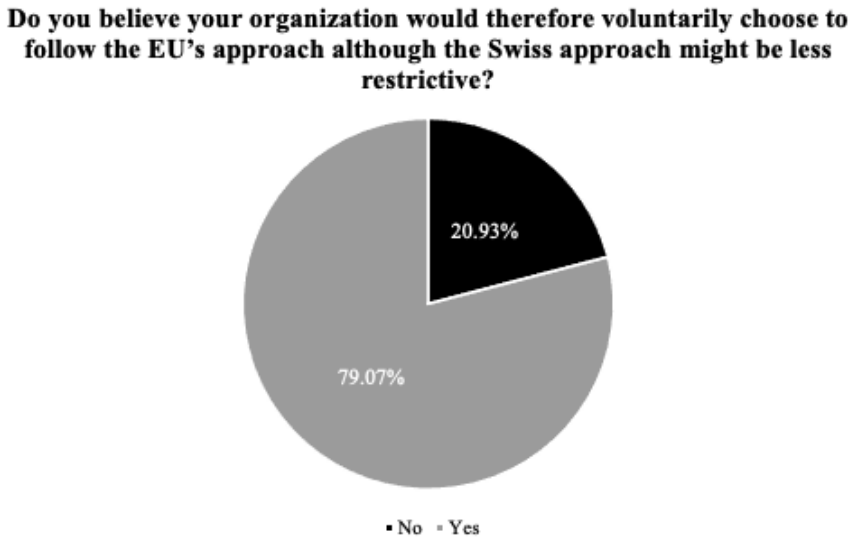


Figure 28: 4 out of 5 companies will comply with the EU AI Act.

The survey results also show that participants whose company does not have a CSR do not have a consensus on whether they would follow the Act. This could be because companies that do not have CSR are probably not that big yet. They might have different plans for the future in terms of expanding to the EU and, therefore, also have differing views on

whether to adopt the EU AI Act or not. On the other side, companies that already have some self-regulation on AI tend to adopt the Act more than companies that do not have self-regulation. This could be due to companies already incorporating the same or similar rules that the EU AI Act covers.

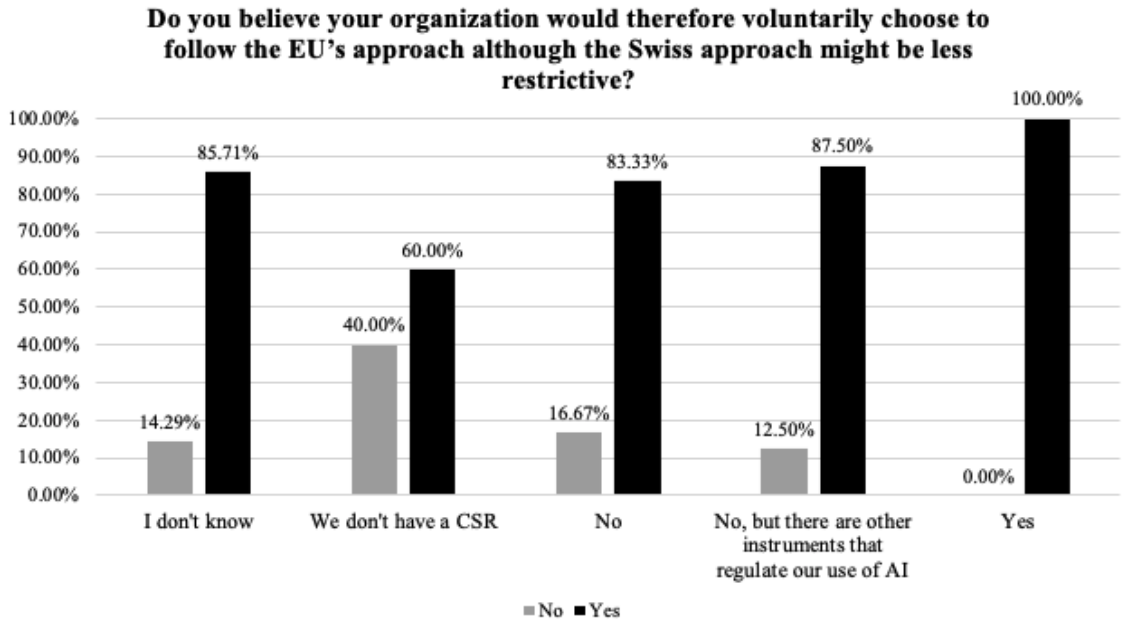


Figure 29: Companies who already self regulate are more likely to also comply with the EU AI Act

9.4 Summary

The survey showed that more than a third of surveyed companies already had to overcome controversial or ethical issues. Such situations ended mostly in projects being abandoned, which is beneficial for responsible AI innovation but could also end up hindering ethically sound and promising projects. Therefore, a more specific AI regulation could serve as a guideline to help avoid these situations. The need of more specific AI regulation was also expressed by the survey sample along with the plan to adopt more AI in the future. This implies that the survey participants are aware of the challenges AI poses, but still do not want to miss out on the benefits it can provide.

While they agreed on needing more specific regulation, the survey participants had quite different opinions about how such regulation should be approached. For example, they almost 50/50 split their opinion on a "responsible AI" label. Many recognize that it could help raise awareness and communicate actions taken against harmful AI to a broader audience. Others doubt its effectiveness since a label for such an intangible technology with a wide range of applications could not cover all of them - or might just be abused as a marketing tool. However, the survey implied that companies looking into using AI favor

a label much more than those already using AI. The sample also differed in which body they favor more to regulate AI.
Business executives trust far more in self-regulation to foster trust among the public than technology executives do. At the same time, one-third of the latter think that trust in AI cannot be built from any regulation. Overall, though, most participants agree that both regulatory bodies can foster trust.
One thing the participants mainly agreed on was compliance with the EU AI Act. Almost four in five participants believed their company would adopt the EU AI Act's rules, even if Swiss AI regulations were less strict.

The survey provided a good insight into the industry's perspective on Swiss AI regulation and showed that although they mostly agree that more specific AI regulation is needed, they differ strongly in what such a regulation should look like. Therefore, the next chapter will have a deeper look at this question by summarizing interviews with AI ethics & regulation experts.

10 Interviews Analysis

After establishing that the industry sample agrees on needing more AI regulation in Switzerland, AI regulation experts were consulted to figure out why AI regulation is needed and how Switzerland should approach it.

10.1 Data Overview

The interview sample consists of six experts that are working in AI Regulation projects or are otherwise frequently confronted with questions about AI regulation and ethics. Table 8 describes the sample.

Industry	**Role**	**Background**
Public Sector	Data Scientist	Technology
Education	Project Member	Legal
Education	Project Member	Legal
Consulting	Ethics Consultant	Socioeconomic
Consulting	Legal Tech Consultant	Legal
IT	Head of Legal	Legal

Table 8: Anonymized description of interviewees

Following the procedure described in the research design outlined in chapter 3.2, the text units of the interviews were assigned a code that summarizes the content of the units. These codes were then grouped into categories. Through this method, 15 categories could be identified. These categories were split among three overarching topics of the interviews: Why is AI regulation necessary, How should AI be regulated, and What is Switzerland's role in AI regulation. Figure 30 displays the identified categories by topic.

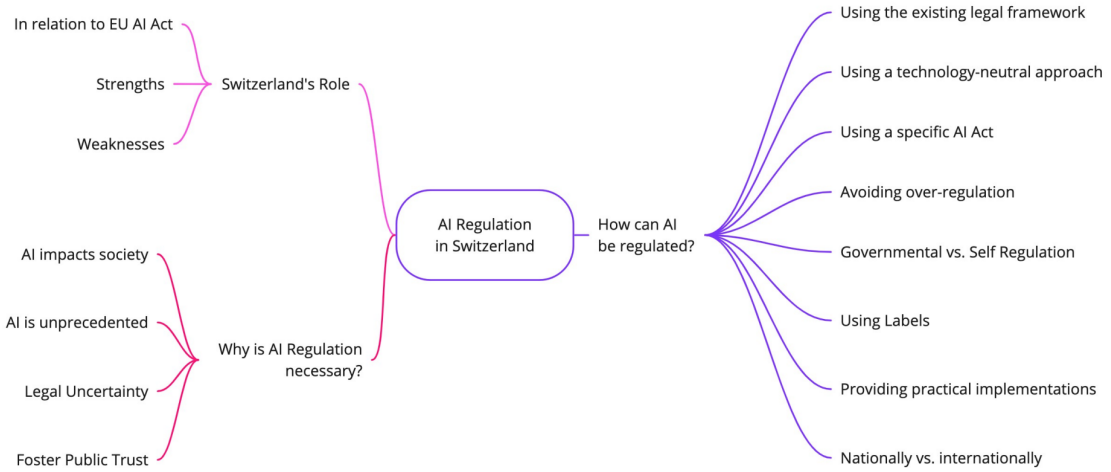


Figure 30: Overview of Categories identified in Interviews

Tables 9 - 11 list the topics and their categories along with the corresponding codes for each topic. The codes are ordered by the number of statements they consist of. A code that has been mentioned more than another is viewed as more relevant. Appendix E displays the categories and codes in a mind map, while appendix F lists all the statements grouped by code.

Category	**Code**
AI impacts society	Generally new challenges for society, Discrimination/Bias, Transparency, Liability, connected challenges, High-Risk-Systems, Manipulation, Autonomy
AI is unprecedented	Term uncertainty, transforms how work is done
Legal Uncertainty	No options for appeal, uncertainty on what is protected, Government needs legal ground to start using AI, prioritization of security standards
Foster Public Trust	Discussions about AI regulation foster trust, currently not much trust in AI

Table 9: Codes and Categories for Topic 1: "Why does AI need Regulation?"

Category	**Code**
Using the existing legal framework	Identify and close gaps, parallel legal structure creates over-regulation
Using a technology-neutral approach	Technology isn't the problem - applications are, can keep up with fast innovation cycles, more tangible, more flexible, allows to cater to sector-specific changes, can lead to legal uncertainty
Using a specific AI Act	Can lead to more coherent understanding, can lead to legal uncertainty, companies might not enter the market
Avoiding over-regulation	Can negatively impact startups and innovation, can lead to extremely high costs
Governmental vs. Self regulation	Governmental regulation needs new perspective, question is about who monitors compliance, self-regulation alone doesn't work, self-regulation is faster, should be a mix in the end, companies who self-regulate shouldn't be punished
Using Labels	Only works if it benefits the company, currently no competition - so no need for a label, provides guidance, companies don't like audits, companies don't like too academic labels, auditing black-box algorithms is difficult
Providing practical implementations	Developers need practical guidance
Nationally vs. internationally	International regulation is extremely difficult, internationally there are different values, general international guidelines can influence national regulation, international regulation would disturb current frameworks

Table 10: Categories and Codes for Topic 2: "How can AI be regulated?"

Category	Code
Strengths	Technology-neutral framework, democratic systems, Data Protection Officer
Weaknesses	Slow democratic process, reserved position, federalism
In relation to the EU AI Act	Shouldn't be applied in Switzerland, companies will have to adopt it anyway, Switzerland could refer to AI Act, comparable to GDPR

Table 11: Categories and Codes for Topic 3: "What is Switzerland's Role?"

The following sections will explain the findings of these categories and codes.

10.2 AI holds unprecedented Challenges for Society - Why AI needs Regulation

Like the survey sample, the interviewees agreed that Switzerland needs more AI regulation. From their statements, four reasons for the necessity of AI regulation could be derived.

Legal uncertainty around AI hurts innovation and fair processes. Three interviewees described that people currently have no options to appeal wrongful decisions influenced or made by an AI algorithm. This is connected to the statements of two interviewees who said that people do not know how they are protected and what options they do have when they experience harm from an AI. The same goes for companies: They often do not know what they need to protect their users or customers from when using AI. In sensitive sectors like health care, legal uncertainty can lead to insufficient security standards. This is why one interviewee argued that legal certainty must be created by setting security standards in AI regulation. This demonstrates that there is quite a large legal uncertainty around Artificial Intelligence, which needs to be addressed to provide safe applications and a fair due process for both companies and the public. Additionally, legal uncertainty holds another risk from a governmental perspective, because government-level institutions need legal grounds that enable them to use AI in the first place. This means that the legal uncertainty surrounding AI actively prevents the government from testing how they could implement AI.

AI is unprecedented. All interviewees discussed that it is hard to define what Artificial Intelligence is since it is quite an abstract concept with many possibilities for application. There is a need to first properly understand it and its challenges before identifying which

means can regulate it effectively. Three interviewees also stated that AI is transforming how work is done and that humans are not the only ones doing the work anymore. For one interviewee, AI is even considered a core technology of their business. As current regulation is not built on such premises, specific AI regulation is needed to include the perspective of machines or algorithms as important business drivers.

Regulation can foster public trust. Three interviewees stated that there is currently not much trust in AI and that many people do not trust AI by default. They agreed that an initiative to implement AI regulation could start discussions about AI problems and AI applications. This educates the public and can build their trust in the technology because they know where it can be implemented and how it will impact them. On the other side, the interviewees mentioned that the specific ways in which the regulation or law is drafted will not substantially impact the public's trust - it is the discussion that makes the difference.

AI impacts society in numerous ways. Almost all interviewees agreed that Artificial Intelligence generally poses new challenges for society that could affect fundamental rights or severe consequences for affected people. For example, a patient could die if an algorithm recommends a treatment that is not fit for that specific patient. In terms of specific challenges, four interviewees mentioned that algorithms could be very biased and create discriminatory outputs. Two of them identified discrimination as one of the most pressing issues in AI. Transparency of algorithms is another issue that many interviewees mentioned. They argued that transparency could serve as a catalyst for solving other issues like discrimination. For example, in more transparent systems developers can identify which factors led to the discriminatory output and correct the model. For the interviewees, it is also crucial that people who come in contact with or are evaluated by AI must know that AI is used, and they should be able to understand the algorithm at hand. This means that the average user must be able to comprehend which inputs led to which outputs. The interviewees also recognized that high-risk systems use AI, such as autonomous weapons or manipulative recommending algorithms. They agreed that these systems pose significant threats and should be highly regulated or banned to minimize abuse. A further issue with AI is that there is no regulation on who holds liability if an AI inflicts harm. The product liability law currently mainly applies to physical products. While there is a consensus that it should also apply to software code, the law does not consider the specific elements of such codes. For example, it assumes that after the product enters the market, it cannot be modified by the manufacturer, which does not apply to software products. This issue is further complicated when considering that AI is not just software code. Therefore, the product liability law does not fully cover AI solutions, contributing to legal uncertainty (see above).

These reasons indicate that current regulation is not suited for handling new challenges coming from AI. Therefore, AI regulation is needed to educate society about AI and

protect it from its dangers. The following section will describe how this can be achieved.

10.3 Extending the existing technology-neutral Framework - How AI should be regulated

The main part of the interviews consisted of discussing what means would be suited to regulate AI in Switzerland and how they should be implemented.

Using the existing legal framework. Most interviewees referenced that there are already many Swiss laws regulating situations in which AI could be applied. For example, the anti-discrimination law applies to everyone, including AI companies. Therefore, the interviewees propose identifying the gaps in those laws that do not cover AI applications and closing them with regulatory actions. Two experts mentioned that building a new AI regulation from the ground up would create over-regulation and confusion because there might be intersections with existing laws.

Using a technology-neutral approach. In three experts' perceptions, AI technology is not causing problems. It's the specific applications that pose challenges. Against this background, they believe that the technology itself should not be regulated, but the applications should. According to the experts, the Swiss traditional technology-neutral approach would be more suited for such a situation. Additionally, the approach provides several benefits: The main benefit is that it can keep up better with rapid innovation cycles. Because it is impossible to guess which technologies will be invented and adopted in the future, it is better to create a regulatory framework that applies to any application of existing or upcoming technologies. Furthermore, a technology-neutral framework can be adapted more flexibly. For example, since the regulatory approach differs for each sector, it is easier to include AI-specific regulations within these existing sector regulations than to create an entirely new law. Experts also stated that a technology-neutral approach is more tangible than an AI Act because there is no abundance of new laws but mostly only adaptions to existing ones. This is easier for the public to understand because they might not be familiar with AI technology. One detriment to a technology-neutral approach is that it might result in a fragmented legal framework down the line. This, in turn, could lead to legal uncertainty because companies struggle to find all the relevant elements for their AI use cases.

Using a specific AI Act. Because many interviewees favor a technology-neutral approach, few endorsed attempting to regulate AI through a specific AI Act.

Those who did mainly explained that an AI Act could "level the playing field" by ensuring that all market players have to follow the same rules. They also stated that a specific AI Act makes it easier for companies to understand the rules they need to follow, as they

are all in one document. This statement contradicts the assumption of experts favoring the technology-neutral approach, who believe an AI Act is too complicated to understand for the public and companies as the EU AI Act shows. One interviewee mentioned that companies already have difficulties finding what regulations apply to their AI use case. It is even more confusing for people who are not quite technologically literate to understand how they are protected through this act. Some interviewees stated that there might also be a risk to using a specific AI Act because companies might just not enter the market if the regulation bans or over-regulates certain types of applications.

Using Labels. The interviewees explained that a label is a good starting point for regulation since it can provide guidance to companies working on an AI solution. However, they stressed that a label can only be effective if it benefits the company. One interviewee explained that since there is currently not much competition around responsible AI in Switzerland, clients do not demand labels, making them not interesting for companies. Another interviewee shared that companies often do not like labels that academic researchers influence too much because they fear it has lost touch with day-to-day business. They also might be afraid of audits as those can bring to light issues that haven't yet been on the company's radar. Additionally, it might be hard to audit algorithms in the first place, because there is not much experience with auditing black-box algorithms.

Avoid over-regulation. Regardless of which form the regulatory approach will take, the interviewees agreed that over-regulation should be avoided. They argued that overly strict regulatory frameworks could lead to high costs for entering the market and hinder startups from developing products. This would be detrimental to AI innovation in Switzerland, as these startups and other companies would give up on projects that might have been beneficial for the public or the companies. Or they might even move their projects to another country where AI is less heavily regulated.

Providing practical implementations. Two interviewees stressed that it is essential to deliver practical implementations as part of a regulatory framework. They explained that developers need practical guidelines to understand the regulation's technical implications. A best practice would be to include questions or criteria for each step of the development and deployment process that translates the overarching rules from the regulation to specific actions within the process.

Self-Regulation vs. governmental regulation. The interviewees did not favor any one specific regulatory body. They rather stated that governmental regulation needs a new perspective and view regulation as a life-cycle instead of a one-off procedure because regulations for rapidly evolving technologies need to be adapted periodically to stay effective. Although the interviewees think self-regulation is better suited to regulate AI since it is faster and more flexible than governmental regulation, they also note that

self-regulation hasn't worked in the grand scheme. One reason might be that not every company is doing it, which can lead to self-regulating companies being "punished" in the market. Therefore, the experts favor a mix of both, governmental and self-regulation, for regulating AI in the long run. In this constellation, the governmental regulation could define the generally applicable rules to level the playing field. Self-regulation can then serve as a "rulebook" that describes how to practically implement the rules within the companies' processes. Furthermore, one interviewee stated that it is not about "who makes the rules" but "who monitors their compliance". This implies that it is not enough to just have a regulatory framework of rules. It also needs to encompass institutions that independently monitor whether regulations are complied with.

National vs. international regulation. The interviewees agree that it is impossible to regulate something internationally, as international political debates about the Covid-19 pandemic and the global climate crisis have shown. One reason for this is because there are different values and understandings of the government's role which complicates discussions about ethics and the scope of regulations. One interviewee questioned whether an international regulation would be useful since it could create chaos by disturbing national legal frameworks that are already in place. However, some interviewees agreed that it's good to have general international guidelines or rules (i.e. from the OECD or the EU AI Act), which influence national regulations. This way, the implemented regulations of member countries generally follow a shared direction and national regulations do not become too fragmented.

This main part of the interview has shown that the experts favor a more technology-neutral approach that takes advantage of existing laws and regulations. However, an efficient regulatory framework would consist of both, governmental and self-regulation, and include practical implementations. They agree that regulation needs to happen on a national level but can be influenced by international guidelines.

10.4 Swiss Options for AI Regulation are restricted - Switzerland's Role

In the discussions with the experts, strengths and weaknesses of Switzerland concerning AI regulation were identified, whereby the experts agreed on the weaknesses more than on the strengths.

Weaknesses. All interviewees mentioned that the democratic process in Switzerland is very slow and therefore one of its most significant weaknesses. Since it often takes more than a year from drafting a regulation to it coming into effect, this process is not suited for regulating technologies subject to short innovation cycles. Another weakness is that

Switzerland usually takes a quite reserved approach to regulation. Four interviewees agreed they would like Switzerland to approach regulation more actively. One interviewee referenced federalism as a weakness in regards to regulation because it could lead to fragmented regulations on a cantonal and a state level.

Strengths. The interviewees had a harder time identifying strengths than weaknesses. Nevertheless, they named three strengths: Two interviewees mentioned that the technology-neutral approach to regulation is an excellent strength of Switzerland because it is more sustainable as it can cover aspects of technologies that are not yet invented. Another interviewee stated that Switzerland has constructed a great regulatory instrument in the form of Data Protection Officers, who - in an extended form - could also be part of an AI regulation. Finally, one interviewee said that the democratic process is certainly not only a weakness but that it does have some benefits too, however they did not mention if they meant in general or for AI regulation.

Switzerland's role concerning the EU. Since Switzerland is surrounded by EU member countries, it stands to reason that many companies also operate in the EU: Three interviewees believe that many Swiss companies will comply with the EU AI Act on their own. Nevertheless, four interviewees believe that Switzerland should not simply adopt the regulations of the EU approach. This is mainly because they prefer the technology-neutral approach and think it fits Swiss culture and democratic processes better. However, two interviewees stated that Switzerland could refer to the EU AI Act in certain situations and adopt some provisions of the Act in its regulations, such as establishing certain institutions. Experts expect Switzerland to react in the same way as with the EU's GDPR directive. Many Swiss companies voluntarily act in a GDPR-compliant manner, and the Swiss Data Protection Act also aligns with the GDPR. The experts' statements and the precedence of GDPR imply that Switzerland cannot just independently define its approach to AI regulation but is implicitly restricted by the EU's approach and rules.

In summary, the interviews have confirmed the survey's findings in that they also describe the need for AI regulation. They have shown that Switzerland is already in a comfortable position regarding AI regulation because of its technology-neutral legal approach. However, there are detriments to such an approach, and some characteristics of the Swiss democratic process make it difficult to keep up with innovation cycles and AI's rapid evolvement.

11 Answers to Research Questions

This chapter combines the theoretic base conducted from literary research with the findings of the survey and interviews in order to answer the research questions.

11.1 What is the current State of AI Regulation in Switzerland?

Like many other countries, Switzerland has acknowledged that AI is a powerful technology with many application scenarios. Nevertheless, it seems to have realized that these applications also challenge society in new ways. The most significant difference in how Switzerland handles these challenges compared to other countries is that Switzerland is much more reactive. It has not defined an AI strategy nor plans to adapt or introduce laws to regulate AI.

This is not to say that Switzerland is entirely inactive. The task force reported in late 2019 that existing laws could handle the new challenges arising from AI. However, the task force stated that the situation would need to be evaluated continuously, given the exponential development of AI. This need for continuous evaluation of AI is also included in the Swiss Digital Strategy. In terms of Artificial Intelligence, the strategy states the goal of providing a framework for transparent and responsible AI applications. However, there are not many practical implications defined yet to achieve this goal. The fact that there is a digital strategy but no specific AI strategy implies that Switzerland prioritizes digital transformation over AI, but it does acknowledge that AI can be a part of digital transformation. This differentiates Switzerland from other regions like the USA, China, and the EU which all have specific AI strategies and regulations drafted.

This difference might lie in Switzerland traditionally favoring technology-neutral approaches to regulation. The approach brings many benefits. For example, it prevents favoring or disapproving of certain technologies and can be more sustainable as it applies to existing and future technologies.

Nevertheless, the government has released guidelines for the internal use of AI solutions. These guidelines contain quite similar aspects to what other countries have mentioned in their AI strategies, for example, developing AI for the public good, ensuring transparency and applicability, and taking part in international discourse.

But not only the Swiss federal government has looked into regulating AI. The canton of Zurich introduced an AI sandbox in early 2022. This sandbox allows testing of cantonal AI applications in a controlled environment to ensure they are safe and fair. Switzerland had been quick to introduce such a sandbox for distributed ledger technology, and it has proven to foster an innovative environment while minimizing harm. Therefore, it is quite

unusual that it took this long to create a sandbox for AI.

In the private sector, Swiss AI regulation consists only of the existing legal framework combined with optional AI-specific self-regulation. This framework seems too leisurely as there is too much legal uncertainty regarding AI for the public and companies. The public has no option to appeal wrongful decisions, and companies do not know for sure whether their AI solution is complying with all applicable rules or if they are missing out on some. This might even lead companies or startups to give up on their projects or move to other countries where they can build their solutions with more legal certainty.

In summary, the Swiss approach is going in the right direction as technology-neutral approaches generally keep up better with technologies that are subject to quick innovation cycles. This is especially helpful when considering Switzerland's slow democratic process. However, there is currently not much public trust in the technology, which might be related to the legal uncertainty surrounding it. This demonstrates that actions need to be taken to foster trust and prevent self-regulating companies from being "punished" in the market.

11.2 What is the Industry's Perspective on AI Regulation?

There is a lot of trust in AI technology from the industry's perspective, as many representatives want to incorporate more of it into their business. This indicates that AI technology has proven to be beneficial for them. However, the industry also seems to have acknowledged the disadvantages and challenges of AI. More than a third of surveyed companies have already encountered a controversial or ethical problem regarding their AI solution. This emphasizes that companies are currently operating with a large amount of legal uncertainty, especially when working with an AI solution that impacts society in a new way. As many companies abandoned their projects to solve controversial or ethical problems, they wasted valuable time and resources. It might be because of experiences with or concerns about such situations that companies overall endorse having more specific regulations for AI.

Experts and industry representatives alike understand that AI regulation cannot be enacted solely by self-regulation. This is evident since not many companies have implemented any form of self-regulation for AI. The lack of self-regulation leads to it being seen as a disadvantage in the market. Most industry representatives agree that governmental regulation - on its own or in combination with self-regulation - should be included in a regulatory framework. They expect the governmental regulation to ensure all market players follow the same set of rules, which could turn self-regulation into more of a benefit by expressing extra effort to possible customers. For example, such extra effort could be communicated by fulfilling the criteria of a "responsible AI" label.
Interestingly, the representatives are very divided in endorsing such a label. They worry

about its effectiveness and question whether it would just be used for marketing purposes instead of ethical considerations. However, companies planning to use AI in the future seem to be quite interested in a label. This implies that companies who are not yet using AI would like to have more guidance on what is "safe" to pursue in the AI industry without worrying about ethical or legal repercussions.

Although they endorse a more specific AI regulation in Switzerland, many companies already expect to comply with the rules of the EU AI Act. This might be due to AI companies wanting to expand to the EU market or already working it. For Switzerland, this means that the regulation should consider the aspects of the EU AI Act similar to GDPR being taken into account with the Swiss Data Privacy Law.

11.3 What are the most important Factors to include in a Regulation Proposition for AI?

To be able to regulate AI, it must be defined what applications are understood as AI applications. Since the AI industry has not been able to define the term “Artificial Intelligence” this proves to be a difficult task. A too loose or too narrow definition could even lead to over- or under-regulation. Therefore, having a definition of AI is an integral part of any AI regulation.

A crucial aspect of AI applications is their opacity. Transparency should be addressed in AI regulation, so AI applications are developed in a way that makes it possible for users to understand how an algorithm came to a particular conclusion. It is also imperative that users know when they are interacting with an AI so they can take action against it if needed.

Discriminating algorithms are at the forefront of negative AI application examples (see chapter 5.1.6). Using biased algorithms can have significant implications for the parties affected, for example, the denial of essential services or the recommendation of an unsuited medical treatment. Therefore, an AI regulation must include measures to avoid biased algorithms.

Besides biased algorithms, experts regard high-risk systems with worry and are torn on whether they would recommend banning them altogether. In any case, they agree that the handling of high-risk systems such as autonomous weapons and general-purpose systems should be addressed in a regulatory framework for AI. Experts said that high-risk systems might be a topic where Switzerland could adopt the EU’s propositions.

Furthermore, AI regulation needs to address the question of liability. Responsibility must not be delegated to machines, but the current product liability law is not suited for digital products. Therefore, it must be addressed who is to be held liable if an algorithm causes

harm.

11.4 What are the Strengths, Weaknesses, Opportunities and Threats for Swiss AI Regulation?

The findings from analysis of the current situation, survey and interviews can be summarized in a SWOT matrix to identify strengths, weaknesses, opportunities and threats for Swiss AI regulation. Figure 31 displays this matrix. A larger version of the matrix can be found in appendix G.

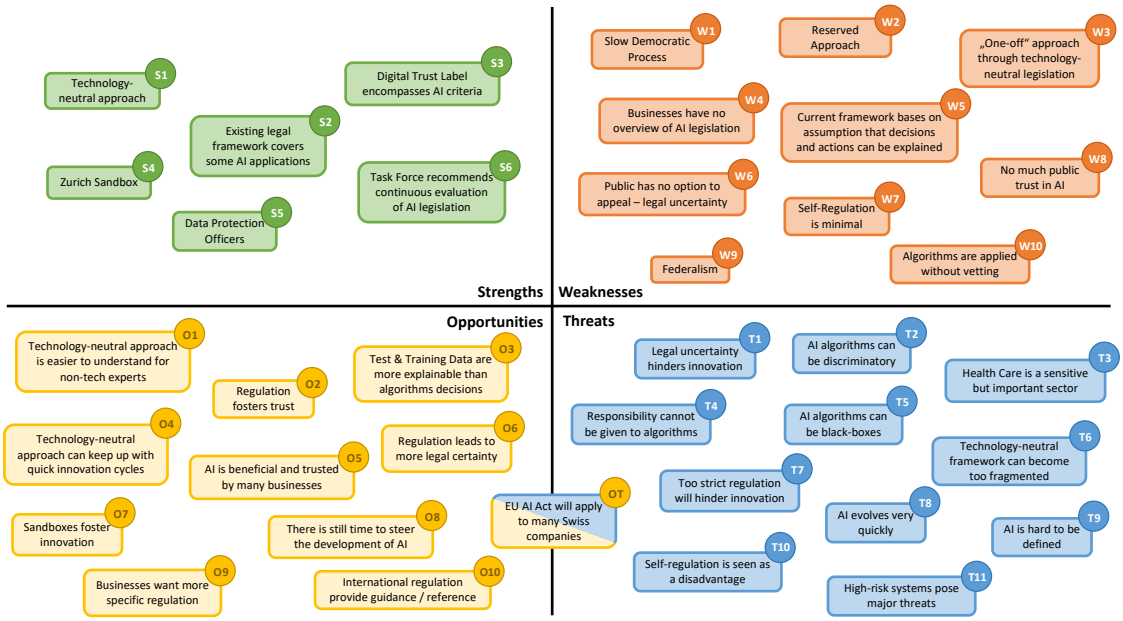


Figure 31: SWOT Matrix of Swiss AI Regulation

The SWOT matrix displays that the Swiss technology-neutral approach can have many benefits. It can keep up better with quick innovation cycles, which works especially well when considering the rapid development of AI and the generally slow democratic process of Swiss legislation (S1, O4, W1, T8). Furthermore, it is hard to define AI in a way that is unanimously agreed on. A technology-neutral approach works without defining what AI is and thus bears less risk of over- or under-regulating based on a too loose or too strict definition (S1, O1, T9). Although the technology-neutral framework applies to all existing and future technologies, the government should listen to the task force's recommendation to continuously evaluate the state of AI regulation. Otherwise, it might miss out on significant technology developments that lead to gaps in the regulation (S1, S6, W3). The continuous adaption and extension of the technology-neutral framework could lead to high fragmentation. Businesses already seem to have trouble identifying

which laws they need to consult for their AI application, which could pose a high risk. A too fragmented framework could become expensive for businesses as they might need to hire lawyers or experts to ensure their AI application is legal. This could hurt innovation as it becomes too expensive. Businesses already call for more specific AI regulations because they want to avoid this scenario (W4, O9, T1, T6).

While the existing legal framework might regulate many AI applications already, for example, through the anti-discrimination law, it still requires that the motivation and reasoning behind decisions and actions can be explained. However, this is not always the case with AI algorithms. Some algorithms are black boxes whose decision processes cannot be explained logically. A manifestation of this threat is that algorithms are applied without being vetted first, for example, the Fotres algorithm, which evaluates prisoners without a scientific basis, and the Ekin Patrol Car, which has facial recognition technology that could lead to mass surveillance (S2, W5, W10, T5). One step to combat the black-box argument in explainable AI is to focus the need for explanation on the test and training data sets as well as the choice of model. These factors have some influence on the algorithm's decision-making process. They can all be explained by the researchers or developers and would be suitable for evaluation with clear criteria (O3, T5).

Switzerland is generally known to hold a more reserved position regarding regulation. However, action must be taken now, as there are already applications of high-risk systems that pose major threats, for example, the Ekin Patrol Car. Experts say there is still time to regulate and thus steer the development of AI, but the government should not wait too long. As Switzerland's neighboring countries, along with the entire EU, will soon have a unified regulation through the EU AI Act, Switzerland must decide how it will react to this approach. As many companies in Switzerland will already have to comply with the EU AI Act due to business interests, Switzerland could use this regulation as a reference for its adaption of AI regulation. Because so many companies will already have to comply with the EU AI Act, Switzerland should also consider these regulations. Otherwise, it might risk its own regulatory framework becoming more complex than the EU AI Act, leading companies and researchers to move abroad for a more stable regulatory situation (W2, O8, O10, T12, OT).

The government cannot just count on self-regulation to regulate AI because many companies in Switzerland do not practice such self-regulation because they see it as a disadvantage in the market. However, the Swiss Digital Trust Label would provide good guidance for AI applications as it also encompasses criteria for trustworthy and responsible AI. As self-regulation is not simply applied, the label's attractiveness should be raised through more communication about the label's advantages to the public and companies (S3, W7, T10). A label is not the only way to strengthen the public's trust in AI, however. Currently, the public does not hold much trust in AI because there is much legal uncertainty

about defending oneself from algorithms. There is no option to appeal against algorithmic decisions. It might even be impossible for some victims to know whether an algorithm or a human made the decision. Governmental regulation could mitigate this opacity and legal uncertainty and thus foster public trust in the technology (W8, O2, O6, W6).

One AI-specific regulatory measure introduced on a governmental level in Switzerland is the AI sandbox, introduced by the canton of Zurich. Such sandboxes have proven to be beneficial for innovation when they were applied to distributed ledger technology. However, the federalism principle in Switzerland carries the risk that if sandboxes are applied in more cantons, these different sandboxes come to different results regarding the safety and fairness of algorithms. Therefore, it must be ensured that the sandbox is implemented in the same way in every canton (S4, O7, W9).

The matrix shows that AI is a beneficial technology and that businesses want to continue applying it. However, they need guidance and clarity on the legal basis as well as more public trust in the technology for their products to succeed in the market. Therefore, the next chapter will focus on describing actionable recommendations to provide clarity and guidance on a governmental and self-regulation level.

12 Recommendations

The thesis has shown that effective AI regulation must come from the governmental level because self-regulation is often seen as a disadvantage in the market and thus not implemented by many companies. However, regulation for such quickly evolving technologies mustn't be seen as a one-off decision. Rather, the regulation affecting AI products and services must be evaluated continuously as the technology evolves. Figure 32 depicts that AI products are not only influenced by governmental and self-regulation but also by (continuously evolving) customer & business needs as well as innovation. As these AI products affect the public in new ways, governmental regulation must continuously be evaluated to discover gaps in the protection of the public.

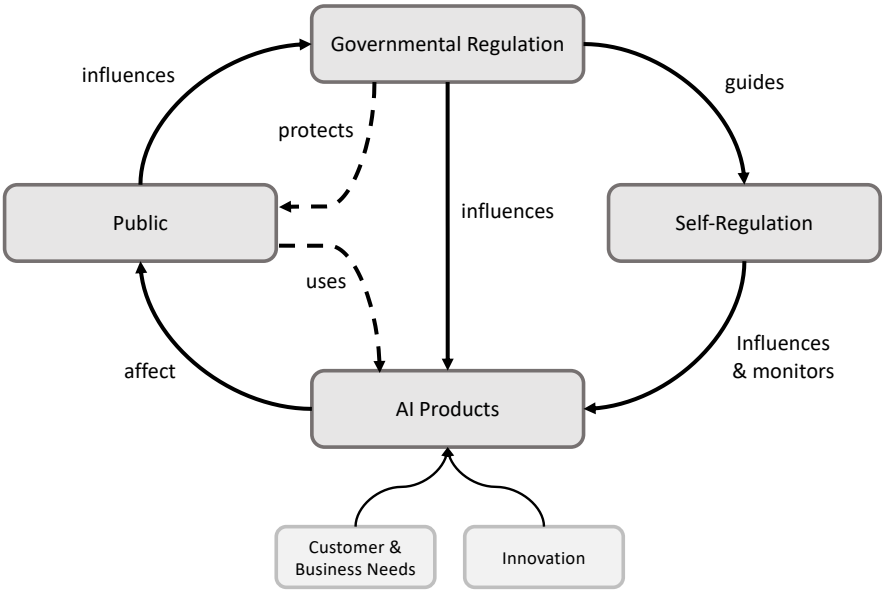


Figure 32: As AI products evolve continuously, so does AI regulation

As governmental and self-regulation both play an important part in influencing AI products and therefore regulate AI technology usage in Switzerland, the following sections will dive deeper into the specific frameworks recommended for governmental and self-regulation, respectively.

12.1 Framework for Governmental Regulation

As Switzerland's technology-neutral approach to regulation is deeply embedded in its culture and legal processes, this approach should also be applied when regulating AI. However, this thesis' research contradicts the government task force's finding that the legal framework needs no adaption. In fact, it has identified ten aspects that should definitely be

included in AI regulation. The aspects can be organized into three categories: Regulate, Inform and Control. Figure 33 depicts these categories and aspects in a governmental regulation framework.

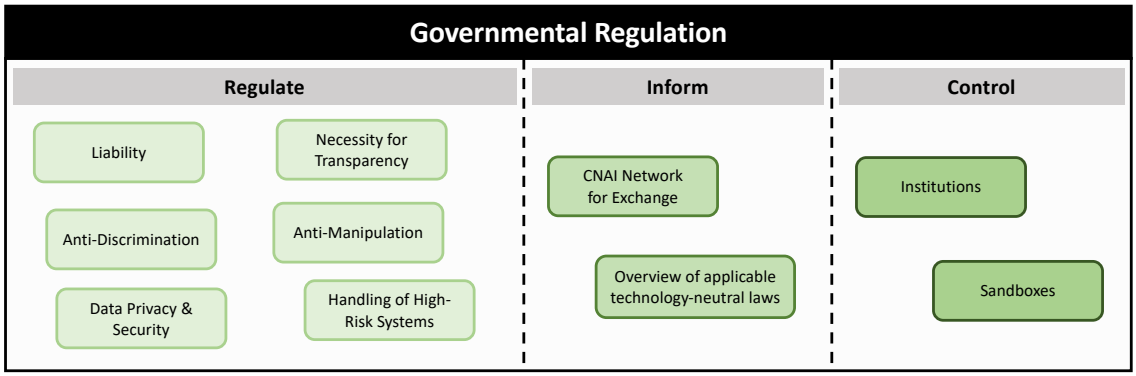


Figure 33: Graphical illustration of a framework for governmental AI regulation

Although Switzerland's technology-neutral approach has many benefits with regard to AI regulation, the following aspects need to be included in a legal framework to effectively regulate AI:

Liability. The question of liability is not fully answered for AI products and services. While there is a consensus that the product liability law applies to software as well, it does not consider the specific properties of such digital products. This also contributes to the legal uncertainty surrounding AI regulation. A clear definition on who is liable during which stage of the (AI) product life-cycle is needed to provide more legal certainty and create a larger sense of responsibility among businesses and people building and using AI products. Specifically, there should be a distinction in scope of liability for companies that provide algorithms and companies that apply those algorithms in their business. This is because companies that are merely using algorithms cannot be held responsible for issues that were created through the training of the algorithm. However, they should be held accountable when they are applying an algorithm in an unfit context or fine-tune the algorithm with biased data. The product liability law needs to be extended with these considerations to truly apply to AI applications.

Necessity for Transparency. One crucial aspect of AI algorithms is the black-box nature of some models. This makes it hard for laws to apply because most of them are based on the assumption that the reason behind certain actions or decisions can be explained. Because of this, Swiss AI regulation needs to manifest that transparency of the algorithm's decision making process is needed for the technology-neutral legal framework to apply. The level to which transparency is required needs to be nuanced depending on the purpose and impact of the algorithm. For example, high risk systems that could change an individual's opinions or way of life considerably (f.e. calculation of risk of recidivism) should require a more detailed level of explanation than lower risk systems that do not have such a strong impact

on the individual (f.e. recommendation of songs in a music app). Also, transparency does not necessarily mean that the algorithm's decision itself needs to be explainable. However, the data sets and data cleaning decisions should always be explainable as well as the reason why the specific model was chosen for the algorithm. These reasons provide at least a basis to interpret which factors influenced the algorithm's decision, which in turn are required when applying the law.

Anti-Discrimination. The research has shown that biased and discriminatory algorithms are very dangerous for society as they can cause a large amount of harm. Switzerland already has an anti-discrimination law in effect, however the burden of proof lies with the discriminated party[144]. The government should evaluate to move the burden of proof from the discriminated to the alleged discriminator as it is especially difficult for victims of discriminatory algorithms to prove misconduct, especially due to the opacity of algorithms and the advanced technical knowledge required to understand them. If the companies held the burden of proof then citizens or users would have more courage to appeal wrongful decisions. Furthermore, the legal framework should ensure that discrimination in algorithms is detected as early as possible to avoid any actual harm. This could be ensured through mandatory testing criteria and reports on these tests when an algorithm processes personal data such as race or gender, among others. Nevertheless, it is not recommended to prohibit using such information in training algorithms as they can be important factors in calculating certain outcomes[8]. Most importantly, companies need to be incentivised by the legal framework to test their algorithm for bias continuously when they are using such data.

Anti-Manipulation. Certain algorithms can have a manipulating effect on people and therefore restrict their fundamental right to autonomy. Especially recommendation systems and image or video altering algorithms (deep fakes) can manipulate public opinion. This endangers democracy as such algorithms spread fake news and polarize people towards one end of the democratic spectrum. As not all algorithms have the same impact on people's autonomy, the government should make it mandatory for possibly manipulative algorithms to first be thoroughly tested in a sandbox environment. This should especially be required of algorithms that produce and alter content (images, videos, text) and that limit individuals' options for neutral information gain, for example through displaying content selectively. As algorithms evolve over time with new data, this testing and admission procedure should be repeated regularly. The government should also evaluate the ban of algorithms with manipulative effects in certain contexts like political marketing or when criminal intent is involved. At the same time, it should foster research of such algorithms so they can be better understood and more easily identified, for example in the detection of deep fakes.

[8]For example, statistics have shown that female convicts are less likely to reoffend, so the gender should be incorporated in calculating the risk of recidivism (see chapter 5.1.2).

Data Privacy & Security. As data is a key aspect of AI algorithms, it must be ensured that the privacy of user data is ensured and that the algorithm is protected from any security threats. Under this aspect, the legal framework should also ensure that users are informed transparently when they are interacting with an AI algorithm and which kind of data the algorithm processes for which purpose. This allows them to make an informed decision about whether they want to continue the interaction or not. These measures will be included in the updated Swiss Data Protection Act (Art. 19 & 21), which is scheduled to come into effect on September 1, 2023[145][146]. However, the act does not mention any consequences beyond fines. This might be a too small incentive for companies as they can plan for violation of the law by creating reserves and they won't lose any progress in training the algorithm if they only have to delete the wrongfully acquired data. A more effective - although quite drastic - consequence might be to include the destruction of algorithms that are trained with this wrongfully acquired data[9].

Handling of High-Risk Systems. Experts debate whether high-risk systems like autonomous weapons or algorithms that can manipulate people should be banned completely. The EU AI Act will ban some of such extremely high-risk systems and introduce clear criteria for others. Given their destructive potential, Swiss AI regulation should address how such systems will be handled in Switzerland. It could be considered to adopt the EU's risk pyramid to identify high-risk systems. If a ban is not deemed reasonable, the government should at least establish a permission testing procedure to vet high-risk systems before they enter the market. This will ensure that all systems in the market are proven to not cause massive harm.

As the research has shown, industry representatives have an interest in more specific AI regulation than the current technology-neutral framework. Interview partners have also expressed concerns that the Swiss AI regulatory framework could become too fragmented to combat legal uncertainty for both companies and the public. It is therefore imperative that the framework includes aspects that address the information and education of affected parties. The AI Competency Network CNAI can already contribute to the education by providing a space where industry professionals can exchange best practices and regulatory aspects to consider. There should also be an overview of the applicable technology-neutral laws for AI applications. This would make it easier for companies to gain an oversight of what they need to comply with and for the public to know how they are protected from algorithmic harm. Such an overview could also be established and published by the Competency Network.

[9]For reference see the case of Weight Watchers: It had trained algorithms with illegally harvested personal and sensitive health data and in March 2022 was forced to delete the data, destroy the algorithm and pay a fine as a result [147]

Regulatory and legal frameworks cannot be effective without controlling compliance. Control mechanisms should therefore also be included in the framework. In the case of AI, it's important that institutions like courts are aware of how the technology-neutral framework applies to AI and what the dangers of the technology are. In addition, an AI-adaption of the established Data Privacy Officers would be the right position with the right skills to enforce compliance. Also, they could be a contact point for complaints about algorithmic harm or questions about AI-relevant legislation. Finally, sandboxes provide an excellent way to test algorithms in a controlled environment and make sure they comply with the legal framework and do not cause harm before allowing them to enter the market.

This framework for governmental regulation shows that there are things the government can improve on to create more legal certainty around AI. However, companies still need to translate these rules into actionable and company-specific measures to comply with the legal framework.

12.2 Framework for Self-Regulation in Companies

Research has shown that regulation is only fully effective if the employees working on AI algorithms know what they practically need to implement into their work processes. Figure 34 depicts a framework for AI self-regulation that can help guide the definition of these practical implications.

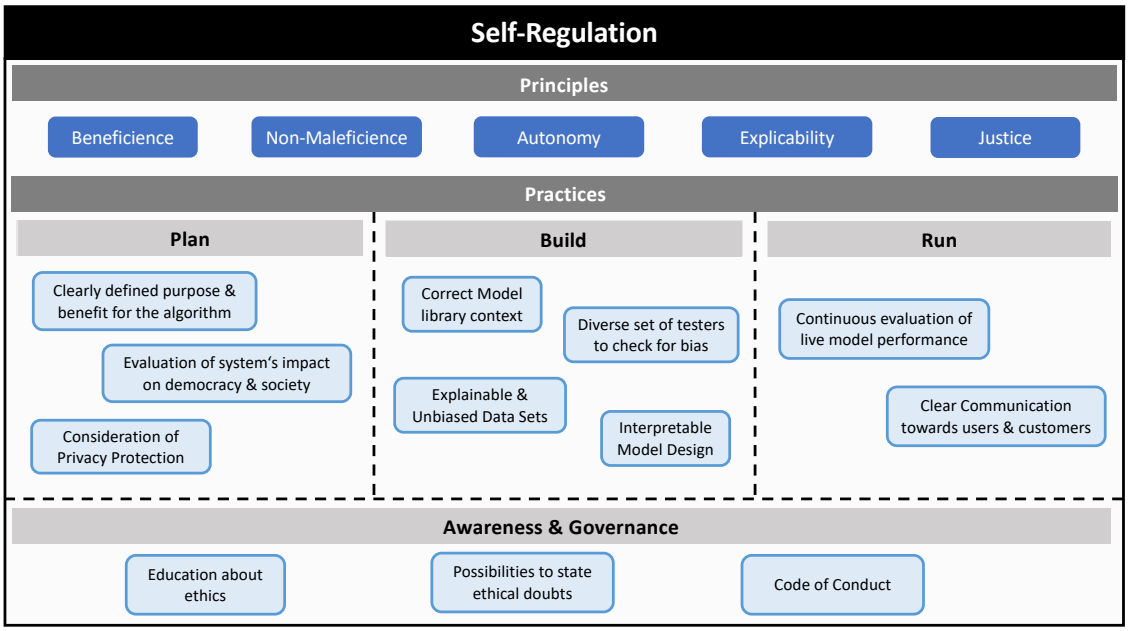


Figure 34: Graphical illustration of a self-regulation framework

The framework consists of two parts - Principles and Practices. The principles aim to guide the strategic decisions and culture of the company. They directly combat the most pressing

AI algorithm challenges, such as discrimination, manipulation, opacity and liability (see table 3 in chapter 5).

The practices aim to translate the ideas of the more strategic principles into tangible and actionable items to consider when working on or with AI products & services. To provide a clear structure, they are divided into the general phases of software development: Plan, Build and Run. Additionally, its integral to any self-regulatory measure that awareness is fostered among employees and that the compliance with the framework is ensured. This is why Awareness & Governance is depicted as a fourth category in the framework.

Plan. During the planning phase of an AI algorithm, it needs to be clearly defined what purpose the algorithm is fulfilling and what benefit this provides for whom. This information is vital to evaluate the algorithm's implications on society and democracy. Clarity on purpose and benefit also is the basis for making an informed decision about the algorithm's compliance with the framework principles. If there are any doubts about the ethical and regulatory soundness of the AI project before the project moves into the building phase, another loop through the planning phase can be taken to define the purpose more clearly and resolve these doubts. As (data) privacy is an essential challenge of AI algorithms, it is also important to consider how an algorithm will respect privacy when it is still in its planning phase because this is the cheapest stage to make adaptions. In later stages, each adaption based on privacy concerns will need more resources.

Build. Developers of AI algorithms have a huge impact on the model's ethical footprint. By considering the original context in the choice of a model, they can prevent many issues. Models should never be applied in a differing context as they might deliver results that are not appropriate for and not sensitive to the new context's characteristics. Another point to consider when choosing a model is that interpretability should always have a high relevance. Thus whenever possible, the simplest model for the purpose of the algorithm should be chosen (Occam's Razor). When training the algorithm, the company should pay close attention to the data sets used. They need to be unbiased and explainable so there is less potential for the algorithm to learn a bias. Finally, bias in algorithms can only be detected if it is tested by a diverse set of people, for example different ethnicities and genders.

Run. Once the algorithm is in production environment, it is important to not just let it run. As algorithms learn and adapt continuously, they need to be evaluated continuously as well. This allows the company to detect any tendencies of the algorithm becoming biased or harmful early enough so it can take measures to correct. In addition, the company must clearly communicate to users and customers when they are interacting with the algorithm. Only then, the user will build trust in the company and their algorithm as well as have the ability to bring any inconsistencies or doubts about results to the attention of the company.

Awareness & Governance. In order for the self-regulatory framework to be truly effective, the company needs to foster ethical awareness amongst employees and govern the compliance with the framework. This can be achieved through education in ethics, especially AI ethics and challenges, and by providing and encouraging the use of anonymous possibilities to speak up about ethical doubts. The self-regulatory measures taken should be manifested in a Code of Conduct and clearly communicated, so that all stakeholders have a coherent understanding of the principles and practices. To govern the compliance with the framework and Code of Conduct, there needs to be someone responsible for compliance and who has the necessary competencies to take action if they notice a breach.

These described practices are inspired by literary research, publicly available ethical frameworks of technology companies like Google or Microsoft and white papers from consulting companies like EY and PWC [148][149][150][151]. The list is not exhaustive. Nevertheless, the framework identifies principles whose practical implication will help avoid legal and ethical AI issues.

Further research would be necessary to identify the most effective and generally applicable practices for Swiss companies. It would possibly also need to be evaluated whether there need to be different levels of practices for different sizes of companies. Until then, companies are encouraged to extend the framework according to their needs.

13 Conclusion

While Artificial Intelligence has already helped build algorithms that are very beneficial to society and companies, there are still many challenges that remain unaccounted for. Specifically, the most impactful challenges are opacity, bias, manipulation, data security and liability. All of these challenges are manifested in examples of harmful AI applications such as deep fakes spreading misinformation or facial recognition algorithms performing significantly worse on non-Caucasian faces. Such examples prove that AI regulation has not yet reached a state where it can prevent AI applications from doing harm.

Although other countries have drafted and put into effect AI regulation to avoid these harmful AI applications, Switzerland remains in a reserved position. Relying on companies to regulate AI themselves is not a viable option, given that self-regulation is rarely applied with moral reasoning and governance. While the Swiss Digital Strategy states the goal of providing a framework for transparent and responsible AI applications, there is not much visible progress towards this goal. A governmental task force has deemed the current technology-neutral framework extensive enough to cover AI related challenges as well. Despite this finding, companies are still operating with a large amount of legal uncertainty. This is a high risk for innovation as it might become too expensive for companies - especially startups - to navigate through the technology-neutral framework in order to figure out what is allowed and what isn't. Swiss companies therefore need guidance and clarity on a legal basis to succeed in the market.

This thesis provides two frameworks - one for governmental regulation and one for self-regulation - that aim to provide guidance and more clarity. The governmental framework recognizes the relevance of a technology-neutral approach in Swiss legislature but also stresses factors that need to be incorporated in Swiss AI regulation to make this approach applicable to challenges posed by AI algorithms. Namely the necessity of a certain level of transparency, measures to avoid bias and definitions on who is liable in the different stages of the application's life cycle are included in the framework. Furthermore, the framework stresses that besides extending the legal framework, the government should also foster information to combat legal uncertainty in companies, and build the required institutions to monitor and enforce compliance.

The governmental framework is a basis upon which companies need to define actionable practices that their employees can apply every day. To support this, the self-regulation framework proposes strategic principles and actionable practices that can be applied in each stage of the software development process such as defining the purpose of the algorithm, selecting the simplest model and continuously checking a running algorithm for bias.

The two frameworks work together with the governmental regulation defining on a high

level what is desired of AI and what is not. Self-regulation then derives an own interpretation of principles that cater to the general regulatory framework and identifies practices that comply with the principles. This setup makes it easier for companies to build a framework that caters to their specific product or service while the government does not need to derive far from its technology-neutral approach. However, it is necessary in both cases to regularly evaluate whether the regulation is still concurrent with the developments in AI technology.

The frameworks serve as inspiration and an entry point for policy makers and industry decision-makers to re-evaluate AI regulation on a governmental and company level, respectively. However, the recommendations and practices depicted are not exhaustive. Further research should be conducted to scientifically identify which practices are most effective to face AI-specific challenges. Since Switzerland is a country of SMEs, it should also be researched if recommended practices differ based on company size.

References

[1] A. Hern, "IBM quits facial-recognition market over police racial-profiling concerns," *The Guardian*, Jun. 9, 2020, ISSN: 0261-3077. [Online]. Available: https://www.theguardian.com/technology/2020/jun/09/ibm-quits-facial-recognition-market-over-law-enforcement-concerns (visited on 11/21/2021).

[2] J. Dastin and M. Vengattil, "Microsoft bans face-recognition sales to police as big tech reacts to protests," *Reuters*, Jun. 11, 2020. [Online]. Available: https://www.reuters.com/article/us-microsoft-facial-recognition-idUSKBN23I2T6 (visited on 11/21/2021).

[3] "OpenAI API," OpenAI. (Jun. 11, 2020), [Online]. Available: https://openai.com/blog/openai-api/ (visited on 11/27/2021).

[4] "Clearview AI | the world's largest facial network," Clearview AI. (), [Online]. Available: https://www.clearview.ai (visited on 11/27/2021).

[5] W. Wang and K. Siau, "Artificial intelligence: A study on governance, policies, and regulations," in *MWAIS 2018 Proceedings*, 2018.

[6] N. A. Smuha, "From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence," *Law, Innovation and Technology*, vol. 13, no. 1, pp. 57–84, 2021.

[7] O. J. Erdélyi and J. Goldsmith, "Regulating artificial intelligence - proposal for a global solution," presented at the AIES, New Orleans, Feb. 2, 2018.

[8] R. Madhavan, J. . Kerr, A. R. Corcos, and B. P. Isaacoff, "Toward trustworthy and responsible artificial intelligence policy development," *IEEE Computer Society*, Oct. 2020.

[9] A. Thierer, A. Castillo O'Sullivan, and R. Russell, *Artificial intelligence and public policy*, 2017.

[10] A. Lauterbach, "Artificial intelligence and policy: Quo vadis?" *DigitalPolicy, REgulation and Governance*, vol. 21, no. 3, pp. 238–263, 2019.

[11] E. Fournier-Tombs, "Towards a united nations internal regulation for artificial intelligence," *Big Data & Society*, vol. 1-5, 2021.

[12] M. Bunz and L. Jancuite, *Artificial intelligence and the internet of things - UK policy opportunities and challenges*, 2018.

[13] M. Guihot, A. Matthew, and N. Suzor, "Nudging robots: Innovative solutions to regulate artificial intelligence.," *Vanderbilt Journal of Entertainment and Technology Law*, vol. 20, no. 2, pp. 385–456, 2017.

[14] H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang, and L. Floridi, "The chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation," *AI & Society*, vol. 36, pp. 59–77, Jun. 17, 2020.

[15] C. I. Gutierrez Gaviria, "The unforeseen consequences of artificial intelligence (AI) on society - a systematic review of regulatory gaps generated by AI in the u.s.," Ph.D. dissertation, Pardee Rand Graduate School, Santa Monica, California, 2020.

[16] N. Braun Binder, T. Burri, M. F. Lohmann, M. Simmler, F. Thouvenin, and K. N. Vokinger, *Künstliche intelligenz: Handlungsbedarf im schweizer recht*, Jun. 28, 2021.

[17] S. Geetha Sethu, "The inevitability of an international regulatory framework for artificial intelligence," presented at the International Conference on Automation, Computational and Technology Management, Dubai: IEEE, 2019.

[18] A. Babuta, M. Oswald, and A. Janjeva, *Artificial intelligence and UK national security - policy considerations*, Apr. 2020.

[19] U. Vesnic-Alujevic, S. Nascimento, and A. Pólvora, "Societal and ethical impacts of artificial intelligence: Critical notes on european policy frameworks," *Telecommunications Policy*, vol. 44, 2020.

[20] N. Döring and J. Bortz, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage, ser. Springer-Lehrbuch. Berlin Heidelberg: Springer, 2016, 1051 pp., ISBN: 978-3-642-41089-5 978-3-642-41088-8.

[21] J. McCarthy, "What is artificial intelligence?" Stanford, CA, Nov. 12, 2007. [Online]. Available: http://jmc.stanford.edu/articles/whatisai.html (visited on 03/19/2022).

[22] S. Legg and M. Hutter, "A collection of definitions of intelligence," *arXiv:0706.3639 [cs]*, Jun. 25, 2007. arXiv: 0706.3639. [Online]. Available: http://arxiv.org/abs/0706.3639 (visited on 04/11/2021).

[23] R. J. Sternberg, "Intelligence," *Dialogues in Clinical Neuroscience*, vol. 14, no. 1, pp. 19–27, Mar. 2012, ISSN: 1294-8322. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341646/ (visited on 02/20/2022).

[24] A. Takacs, I. Rudas, D. Bosl, and T. Haidegger, "Highly automated vehicles and self-driving cars [industry tutorial]," *IEEE Robotics & Automation Magazine*, vol. 25, pp. 106–112, Dec. 1, 2018. DOI: 10.1109/MRA.2018.2874301.

[25] IBM cloud eduction. "What is artificial intelligence (AI)? | IBM." (2021), [Online]. Available: https://www.ibm.com/cloud/learn/what-is-artificial-intelligence (visited on 02/20/2022).

[26] M. Mitchell, *Why AI is harder than we think*, Apr. 26, 2021. [Online]. Available: https://arxiv.org/abs/2104.12871v2.

[27] J. Flowers, "Strong and weak AI: Deweyan considerations," in *AAAI Spring Symposium: Towards Conscious AI Systems*, 2019.

[28] OpenAI. "Better language models and their implications," OpenAI. (Feb. 14, 2019), [Online]. Available: https://openai.com/blog/better-language-models/ (visited on 03/21/2022).

[29] InsightSoftware. "Machine learning vs. traditional programming," insightsoftware. (Feb. 15, 2019), [Online]. Available: https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/ (visited on 05/17/2022).

[30] V. Rajaraman, "JohnMcCarthy — father of artificial intelligence," *Resonance*, vol. 19, no. 3, pp. 198–207, Mar. 1, 2014, ISSN: 0973-712X. DOI: 10.1007/s12045-014-0027-9. [Online]. Available: https://doi.org/10.1007/s12045-014-0027-9 (visited on 03/19/2021).

[31] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," *AI Magazine*, vol. 27, no. 4, pp. 12–12, Dec. 15, 2006, Number: 4, ISSN: 2371-9621. DOI: 10.1609/aimag.v27i4.1904. [Online]. Available: https://ojs.aaai.org/index.php/aimagazine/article/view/1904 (visited on 03/19/2022).

[32] W. Ertel, *Introduction to Artificial Intelligence*, ser. Undergraduate Topics in Computer Science. Cham: Springer International Publishing, 2017, ISBN: 978-3-319-58486-7 978-3-319-58487-4. DOI: 10.1007/978-3-319-58487-4. [Online]. Available: http://link.springer.com/10.1007/978-3-319-58487-4 (visited on 03/19/2021).

[33] E. Mohamed, "The relation of artificial intelligence with internet of things: A survey," *Journal of Cybersecurity and Information Management (JCIM*, vol. Vol. 1, no. 1, pp. 30–34, Jan. 1, 2020. [Online]. Available: https://www.researchgate.net/publication/340006839_The_Relation_Of_Artificial_Intelligence_With_Internet_Of_Things_A_survey.

[34] Towards Data Science. "A concise history of neural networks," Medium. (Apr. 8, 2019), [Online]. Available: https://towardsdatascience.com/a-concise-history-of-neural-networks-2070655d3fec (visited on 04/02/2022).

[35] J. Lee, "Introduction: The development and application of AI technology," in *Industrial AI: Applications with Sustainable Performance*, J. Lee, Ed., Singapore: Springer, 2020, pp. 1–4, ISBN: 9789811521447. DOI: 10.1007/978-981-15-2144-7_1. [Online]. Available: https://doi.org/10.1007/978-981-15-2144-7_1 (visited on 03/02/2022).

[36] Actuaries Digital. "History of AI winters - history of AI winters | actuaries digital." (), [Online]. Available: https://www.actuaries.digital/2018/09/05/history-of-ai-winters/ (visited on 03/21/2022).

[37] "3 trends surface in the gartner emerging technologies hype cycle for 2021," Gartner. (), [Online]. Available: https://www.gartner.com/smarterwithgartner/3-themes-surface-in-the-2021-hype-cycle-for-emerging-technologies (visited on 03/21/2022).

[38] T. Reitmaier, *Aktives Lernen für Klassifikationsprobleme unter der Nutzung von Strukturinformationen*. kassel university press GmbH, Oct. 5, 2015, 251 pp., Google-Books-ID: 6UioCgAAQBAJ, ISBN: 978-3-86219-998-3.

[39] L. Pierson, *Data Science für Dummies*. John Wiley & Sons, Apr. 22, 2016, 446 pp., Google-Books-ID: bqIIDAAAQBAJ, ISBN: 978-3-527-80675-1.

[40] Cognub. "COGNITIVE COMPUTING AND MACHINE LEARNING," Cognub. (), [Online]. Available: http://www.cognub.com/index.php/cognitive-platform/ (visited on 03/19/2021).

[41] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 17, 2015. [Online]. Available: https://www.science.org/doi/full/10.1126/science.aaa8415 (visited on 03/19/2022).

[42] K. Krzyk. "Coding deep learning for beginners," Medium. (Dec. 4, 2018), [Online]. Available: https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d (visited on 03/12/2021).

[43] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *Journal of Machine Learning Research*, vol. 12, no. 29, pp. 1069–1109, 2011. [Online]. Available: http://jmlr.org/papers/v12/chaudhuri11a.html.

[44] G. Forkuor, O. K. L. Hounkpatin, G. Welp, and M. Thiel, "High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear regression models," *PLOS ONE*, vol. 12, no. 1, e0170478, Jan. 23, 2017, Publisher: Public Library of Science, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0170478. [Online]. Available:

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170478 (visited on 03/12/2021).

[45] "Überwachtes & unüberwachtes Lernen im Machine Learning," divisio. (May 24, 2019), [Online]. Available: https://divis.io/2019/05/ki-leicht-erklaert-teil-5-ueberwachtes-unueberwachtes-lernen-ml/ (visited on 05/09/2021).

[46] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: An overview," *Journal of Physics: Conference Series*, vol. 1142, p. 012 012, Nov. 1, 2018. DOI: 10.1088/1742-6596/1142/1/012012.

[47] L. Deng and X. Li, "Machine learning paradigms for speech recognition: An overview," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 1060–1089, May 2013, Conference Name: IEEE Transactions on Audio, Speech, and Language Processing, ISSN: 1558-7924. DOI: 10.1109/TASL.2013.2244083.

[48] V. Kumar, D. Kalitin, and P. Tiwari, "Unsupervised learning dimensionality reduction algorithm PCA for face recognition," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, May 2017, pp. 32–37. DOI: 10.1109/CCAA.2017.8229826.

[49] L. Wuttke. "Unsupervised Learning: Definition, Arten & Beispiele - datasolut Wiki," datasolut GmbH. (), [Online]. Available: https://datasolut.com/wiki/unsupervised-learning/ (visited on 03/10/2021).

[50] C. Szepesvári, "Algorithms for reinforcement learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 4, no. 1, pp. 1–103, Jan. 1, 2010, Publisher: Morgan & Claypool Publishers, ISSN: 1939-4608. DOI: 10.2200/S00268ED1V01Y201005AIM009. [Online]. Available: https://www.morganclaypool.com/doi/abs/10.2200/S00268ED1V01Y201005AIM009 (visited on 03/10/2022).

[51] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*, Second edition, ser. Adaptive computation and machine learning series. Cambridge, Massachusetts: The MIT Press, 2018, 526 pp., ISBN: 978-0-262-03924-6.

[52] X. L. Dong and T. Rekatsinas, "Data integration and machine learning: A natural synergy," in *Proceedings of the 2018 International Conference on Management of Data*, ser. SIGMOD '18, New York, NY, USA: Association for Computing Machinery, May 27, 2018, pp. 1645–1650, ISBN: 978-1-4503-4703-7. DOI: 10.1145/3183713.3197387. [Online]. Available: https://doi.org/10.1145/3183713.3197387 (visited on 03/19/2022).

[53] N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data management challenges in production machine learning," in *Proceedings of the 2017 ACM International Conference on Management of Data*, ser. SIGMOD '17, New York, NY, USA: Association for Computing Machinery, May 9, 2017, pp. 1723–1726, ISBN: 978-1-4503-4197-4. DOI: 10.1145/3035918.3054782. [Online]. Available: https://doi.org/10.1145/3035918.3054782 (visited on 03/21/2022).

[54] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, Number: 7553 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/nature14539. [Online]. Available: https://www.nature.com/articles/nature14539 (visited on 03/15/2022).

[55] QuantDare. "What is the difference between deep learning and machine learning?" (Aug. 1, 2019), [Online]. Available: https://quantdare.com/what-is-the-difference-between-deep-learning-and-machine-learning/ (visited on 03/16/2022).

[56] S. Sharma. "Activation functions in neural networks," Medium. (Jul. 4, 2021), [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6 (visited on 04/02/2022).

[57] S. Rao. "The benefits of AI in construction." (), [Online]. Available: https://constructible.trimble.com/construction-industry/the-benefits-of-ai-in-construction (visited on 03/20/2022).

[58] Beyondminds. "AI-based predictive maintenance for manufacturing," BeyondMinds. (Jun. 20, 2021), [Online]. Available: https://beyondminds.ai/blog/predictive-maintenance-using-ai-to-take-manufacturing-a-big-step-forward/ (visited on 03/20/2022).

[59] S. Rao. "48% of consumer goods companies turn to AI and automation — the research tells us why," The 360 Blog from Salesforce. (Apr. 30, 2021), [Online]. Available: https://www.salesforce.com/blog/consumer-goods-companies/ (visited on 03/20/2022).

[60] H. Matyjaszek. "10 remarkable applications of AI in the energy & utilities industry," Energy Live News. (Dec. 8, 2021), [Online]. Available: https://www.energylivenews.com/2021/12/08/10-remarkable-applications-of-ai-in-the-energy-utilities-industry/ (visited on 03/20/2022).

[61] A. Schroer. "20 key examples of AI in finance you should know 2022 | built in," Built In. (Jul. 30, 2021), [Online]. Available: https://builtin.com/artificial-intelligence/ai-finance-banking-applications-companies (visited on 03/20/2022).

[62] C. Dominish. "How AI is changing the property industry forever," Digital Nation. (), [Online]. Available: https://www.itnews.com.au/digitalnation/news/how-ai-is-changing-the-property-industry-forever-576623 (visited on 03/20/2022).

[63] S. Daley. "35 AI in healthcare examples you should know | built in." (Aug. 12, 2021), [Online]. Available: https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare (visited on 03/20/2022).

[64] A. Tzachor, M. Devare, B. King, S. Avin, and S. Ó hÉigeartaigh, "Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 104–109, Feb. 2022, Number: 2 Publisher: Nature Publishing Group, ISSN: 2522-5839. DOI: 10.1038/s42256-022-00440-4. [Online]. Available: https://www.nature.com/articles/s42256-022-00440-4 (visited on 03/20/2022).

[65] D. Coldewey. "DeepMind's AlphaCode AI writes code at a competitive level," TechCrunch. (Feb. 2, 2022), [Online]. Available: https://social.techcrunch.com/2022/02/02/deepminds-alphacode-ai-writes-code-at-a-competitive-level/ (visited on 03/20/2022).

[66] A. Tsamados, N. Aggarwal, J. Cowls, *et al.*, "The ethics of algorithms: Key problems and solutions," *AI & SOCIETY*, vol. 37, no. 1, pp. 215–230, Mar. 1, 2022, ISSN: 1435-5655. DOI: 10.1007/s00146-021-01154-8. [Online]. Available: https://doi.org/10.1007/s00146-021-01154-8 (visited on 03/12/2022).

[67] W. Knight. "An AI that writes convincing prose risks mass-producing fake news," MIT Technology Review. (2019), [Online]. Available: https://www.technologyreview.com/2019/02/14/137426/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish/ (visited on 03/13/2022).

[68] A. Banerjea. "Digital war: How russia is using deep fakes in ukraine for propaganda," Business Today. (2022), [Online]. Available: https://www.businesstoday.in/latest/world/story/digital-war-how-russia-is-using-deep-fakes-in-ukraine-for-propaganda-324531-2022-03-02 (visited on 03/13/2022).

[69] A. C. Madrigal. "What facebook did to american democracy," The Atlantic. Section: Technology. (Oct. 12, 2017), [Online]. Available: https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/ (visited on 03/12/2022).

[70] M. Scott. "As war in ukraine evolves, so do disinformation tactics," POLITICO. (Mar. 10, 2022), [Online]. Available: https://www.politico.eu/article/ukraine-russia-disinformation-propaganda/ (visited on 03/12/2022).

[71] B. Marr. "How much data do we create every day? the mind-blowing stats everyone should read," Forbes. Section: Enterprise & Cloud. (), [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/ (visited on 03/13/2022).

[72] T. Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," *Minds and Machines*, vol. 30, no. 1, pp. 99–120, Mar. 1, 2020, ISSN: 1572-8641. DOI: 10.1007/s11023-020-09517-8. [Online]. Available: https://doi.org/10.1007/s11023-020-09517-8 (visited on 11/27/2021).

[73] Nationales Zentrum für Cybersicherheit NCSC, *Halbjahresbericht 2021/i (januar - juni)*, 2021. (visited on 03/13/2022).

[74] "What is GDPR, the EU's new data protection law?" GDPR.eu. Section: GDPR Overview. (Nov. 7, 2018), [Online]. Available: https://gdpr.eu/what-is-gdpr/ (visited on 03/13/2022).

[75] L. Bühlmann and M. Reinle. "Neues Schweizer Datenschutzrecht: wichtigste Regelungen der DSG-Revision im Überblick," MLL News Portal. (Oct. 19, 2020), [Online]. Available: https://www.mll-news.com/neues-schweizer-datenschutzrecht-wichtigste-regelungen-der-dsg-revision-im-ueberblick/ (visited on 03/13/2022).

[76] A. Gold. "New china privacy law leaves u.s. behind — and tech companies confused," Axios. (Nov. 23, 2021), [Online]. Available: https://www.axios.com/china-privacy-law-leaves-us-behind-31875e27-e25a-4322-b262-3878445f4fa9.html (visited on 03/13/2022).

[77] SRF. "Rückfallrisiko bei Straftätern - Die grosse Screening-Maschine," Schweizer Radio und Fernsehen (SRF). Section: News. (Jun. 18, 2018), [Online]. Available: https://www.srf.ch/news/schweiz/rueckfallrisiko-bei-straftaetern-die-grosse-screening-maschine (visited on 11/21/2021).

[78] M. Sprecher, "Deliktorientierte Psychotherapie: Meinen Kopf kriegt ihr nicht!" *Die Zeit*, Mar. 5, 2018, ISSN: 0044-2070. [Online]. Available: https://www.zeit.de/2018/10/deliktorientierte-psychotherapie-gefaengnis-zuchthaus-strafe-freiheit/seite-3 (visited on 11/29/2021).

[79] F. Keller, A. Kliemann, D. Karanedialkova, *et al.*, "Beurteilerübereinstimmung im Forensischen Operationalisierten Therapie-Risiko-Evaluations-System," *Nervenheilkunde*, vol. 30, no. 10, pp. 813–817, 2011, Publisher: Schattauer GmbH, ISSN: 0722-1541, 2567-5788. DOI: 10.1055/s-0038-1628429. [Online]. Available: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0038-1628429 (visited on 11/29/2021).

[80] Neue Zürcher Zeitung, "Das prognoseinstrument zur abschätzung der rückfallgefahr von delinquenten stösst auf kritik," *Neue Züricher Zeitung Online*, Sep. 23, 2020. [Online]. Available: https://www.nzz.ch/schweiz/kritik-an-prognose-instrument-zur-abschaetzung-der-rueckfallgefahr-von-delinquenten-ld.1576710?reduced=true (visited on 11/21/2021).

[81] AlgorithmWatch Schweiz. "Minuspunkte für Hautfarbe," AlgorithmWatch Schweiz. (Dec. 3, 2020), [Online]. Available: https://algorithmwatch.ch/de/racial-health-bias/ (visited on 11/20/2021).

[82] D. A. Vyas, L. G. Eisenstein, and D. S. Jones, "Hidden in plain sight — reconsidering the use of race correction in clinical algorithms," *The new england journal of medicine*, p. 9, 2020.

[83] N. R. Powe, "Black kidney function matters: Use or misuse of race?" *JAMA*, vol. 324, no. 8, pp. 737–738, Aug. 25, 2020, ISSN: 0098-7484. DOI: 10.1001/jama.2020.13378. [Online]. Available: https://doi.org/10.1001/jama.2020.13378 (visited on 11/29/2021).

[84] N. D. Eneanya, W. Yang, and P. P. Reese, "Reconsidering the consequences of using race to estimate kidney function," *JAMA*, vol. 322, no. 2, pp. 113–114, Jul. 9, 2019, ISSN: 0098-7484. DOI: 10.1001/jama.2019.5774. [Online]. Available: https://doi.org/10.1001/jama.2019.5774 (visited on 11/29/2021).

[85] J. A. Diao, G. J. Wu, H. A. Taylor, *et al.*, "Clinical implications of removing race from estimates of kidney function," *JAMA*, vol. 325, no. 2, pp. 184–186, Jan. 12, 2021, ISSN: 0098-7484. DOI: 10.1001/jama.2020.22124. [Online]. Available: https://doi.org/10.1001/jama.2020.22124 (visited on 11/29/2021).

[86] J. Madhusoodanan, "Is a racially-biased algorithm delaying health care for one million black people?" *Nature*, vol. 588, no. 7839, pp. 546–547, Dec. 16, 2020, Bandiera_abtest: a Cg_type: News Number: 7839 Publisher: Nature Publishing Group Subject_term: Society, Public health, Health care, Diseases, Biochemistry. DOI: 10.1038/d41586-020-03419-6. [Online]. Available: https://www.nature.com/articles/d41586-020-03419-6 (visited on 11/29/2021).

[87] Tagesanzeiger. "Umstrittene Technologie im Einsatz – So jagen Schweizer Polizisten mit Gesichtserkennung Verbrecher," Tages-Anzeiger. (Mar. 13, 2021), [Online]. Available: https://www.tagesanzeiger.ch/so-jagen-schweizer-polizisten-mit-gesichtserkennung-verbrecher-608167461846 (visited on 11/21/2021).

[88] D. Castelvecchi, "Is facial recognition too biased to be let loose?" *Nature*, vol. 587, no. 7834, pp. 347–349, Nov. 18, 2020, Bandiera_abtest: a Cg_type: News Feature Number: 7834 Publisher: Nature Publishing Group Subject_term: Machine learning, Ethics, Computer science. DOI: 10.1038/d41586-020-03186-4. [Online]. Available: https://www.nature.com/articles/d41586-020-03186-4 (visited on 11/21/2021).

[89] A. Fichter. "Jedes Gesicht in Sekunden identifiziert," Republik. (2019), [Online]. Available: https://www.republik.ch/2019/10/29/360-ueberwachung-made-in-turkey-jedes-gesicht-in-sekunden-identifiziert (visited on 03/13/2022).

[90] BBC News, "Facebook apology as AI labels black men 'primates'," *BBC News*, Sep. 6, 2021. [Online]. Available: https://www.bbc.com/news/technology-58462511 (visited on 11/21/2021).

[91] R. Mac, "Facebook apologizes after a.i. puts 'primates' label on video of black men," *The New York Times*, Sep. 3, 2021, ISSN: 0362-4331. [Online]. Available: https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html (visited on 11/21/2021).

[92] The Washington Post, "Facebook is ending use of facial recognition software, deleting data on more than a billion people," *Washington Post*, Nov. 2, 2021, ISSN: 0190-8286. [Online]. Available: https://www.washingtonpost.com/technology/2021/11/02/facebook-ends-facial-recognition/ (visited on 11/21/2021).

[93] AlgorithmWatch Schweiz. "Image classification algorithms at apple, google still push racist tropes," AlgorithmWatch. (May 14, 2021), [Online]. Available: https://algorithmwatch.org/en/apple-google-computer-vision-racist/ (visited on 11/21/2021).

[94] ——, "How french welfare services are creating 'robo-debt'," AlgorithmWatch. (Apr. 15, 2021), [Online]. Available: https://algorithmwatch.org/en/robo-debt-france/ (visited on 11/21/2021).

[95] ——, "Schweden: Fehlerhafter Algorithmus stoppt Zahlungen für über 70.000 Arbeitslose," AlgorithmWatch. (Feb. 28, 2019), [Online]. Available: https://algorithmwatch.org/de/rogue-algorithm-in-sweden-stops-welfare-payments/ (visited on 11/21/2021).

[96] ——, "Spanien: Rechtsstreit um den Code eines Algorithmus," AlgorithmWatch. (Aug. 12, 2019), [Online]. Available: https://algorithmwatch.org/de/spanien-rechtsstreit-um-den-code-eines-algorithmus/ (visited on 11/21/2021).

[97] The Washington Post, "A face-scanning algorithm increasingly decides whether you deserve the job," *Washington Post*, Nov. 6, 2019, ISSN: 0190-8286. [Online]. Available: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/ (visited on 11/21/2021).

[98] C. Meier. "Wenn Algorithmen rekrutieren – und diskriminieren," elleXX. (Oct. 29, 2021), [Online]. Available: https://www.ellexx.com/de/themen/karriere/wenn-algorithmen-rekrutieren-und-diskriminieren/ (visited on 11/21/2021).

[99] S. Burgess. "Ukraine war: Deepfake video of zelenskyy telling ukrainians to 'lay down arms' debunked," Sky News. (Mar. 17, 2022), [Online]. Available: https://news.sky.com/story/ukraine-war-deepfake-video-of-zelenskyy-telling-ukrainians-to-lay-down-arms-debunked-12567789 (visited on 03/17/2022).

[100] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices," *Science and Engineering Ethics*, vol. 26, no. 4, pp. 2141–2168, Aug. 1, 2020, ISSN: 1471-5546. DOI: 10.1007/s11948-019-00165-5. [Online]. Available: https://doi.org/10.1007/s11948-019-00165-5 (visited on 11/27/2021).

[101] A. McNamara, J. Smith, and E. Murphy-Hill, "Does ACM's code of ethics change ethical decision making in software development?" In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ser. ESEC/FSE 2018, New York, NY, USA: Association for Computing Machinery, Oct. 26, 2018, pp. 729–733, ISBN: 978-1-4503-5573-5. DOI: 10.1145/3236024.3264833. [Online]. Available: https://doi.org/10.1145/3236024.3264833 (visited on 03/12/2022).

[102] P. Boddington, *Towards a code of ethics for artificial intelligence*. Cham: Springer, 2017.

[103] B. Orbach, "What is regulation?" Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2143385, Sep. 7, 2012. [Online]. Available: https://papers.ssrn.com/abstract=2143385 (visited on 03/14/2022).

[104] Cambridge Dictionary. "Regulation," Cambridge Dictionary. (), [Online]. Available: https://dictionary.cambridge.org/dictionary/english/regulation (visited on 03/14/2022).

[105] R. Steurer, "Disentangling governance: A synoptic view of regulation by government, business and civil society," *Policy Sciences*, vol. 46, no. 4, pp. 387–410, Dec. 1, 2013, ISSN: 1573-0891. DOI: 10.1007/s11077-013-9177-y. [Online]. Available: https://doi.org/10.1007/s11077-013-9177-y (visited on 11/22/2021).

[106] Partnership On AI. "Partner index," Partnership on AI. (), [Online]. Available: https://partnershiponai.org/partners/ (visited on 03/14/2022).

[107] N. Statt. "Google reportedly leaving project maven military AI program after 2019," The Verge. (Jun. 1, 2018), [Online]. Available: https://www.theverge.com/2018/6/1/17418406/google-maven-drone-imagery-ai-contract-expire (visited on 03/14/2022).

[108] C. Lecher. "The employee letter denouncing microsoft's ICE contract now has over 300 signatures," The Verge. (Jun. 21, 2018), [Online]. Available: https://www.theverge.com/2018/6/21/17488328/microsoft-ice-employees-signatures-protest (visited on 03/14/2022).

[109] M. Fenwick, W. A. Kaal, and E. P. M. Vermeulen, "Regulation tomorrow: What happens when technology is faster than the law?" Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2834531, 2017. DOI: 10.2139/ssrn.2834531. [Online]. Available: https://papers.ssrn.com/abstract=2834531 (visited on 03/14/2022).

[110] Future of Life Institute. "National and international AI strategies," Future of Life Institute. (), [Online]. Available: https://futureoflife.org/national-international-ai-strategies/ (visited on 11/01/2021).

[111] OECD.AI. "Policy initiatives." (), [Online]. Available: https://oecd.ai/en/policy-initiatives (visited on 11/02/2021).

[112] ——, "OECD AI's live repository of over 260 AI strategies & policies." (), [Online]. Available: https://oecd.ai/en/dashboards (visited on 11/05/2021).

[113] Oxford Insights, *Government AI readiness index 2020*, 2020.

[114] Government of the Republic of Estonia, *Estonia's national AI strategy*, Jul. 2019.

[115] Norwegian Ministry of Local Government {and} Modernisation, *National strategy for artificial intelligence*.

[116] The Government of the grand Duchy of Luxembourg, *Artificial intelligence: A strategic vision for luxembourg*.

[117] Government Offices of Sweden, *National approach to artificial intelligence*, 2018.

[118] National Security Commission on Artificial Intelligence, *Final report*.

[119] "Wyden, booker and clarke introduce algorithmic accountability act of 2022 to require new transparency and accountability for automated decision systems | u.s. senator ron wyden of oregon," U.S. Senator Ron Wyden of Oregon. (Feb. 3, 2022), [Online]. Available: https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems (visited on 04/03/2022).

[120] HM Government, *National AI strategy*, Sep. 2021.

[121] The Federal Government, *Artificial intelligence strategy of the german federal government*, Dec. 2020.

[122] J. Conrad, "China is about to regulate AI—and the world is watching," *Wired*, Feb. 22, 2022, Section: tags, ISSN: 1059-1028. [Online]. Available: https://www.wired.com/story/china-regulate-ai-world-watching/ (visited on 04/03/2022).

[123] State Council of China, *A next generation artificial intelligence development plan*, trans. by Leiden Asia Centre, Yale Law School Paul Tsai China Center, Eurasia Group, and E. Kania, Jul. 2017.

[124] 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*, Apr. 2021.

[125] SERI, *Herausforderungen der künstlichen intelligenz - bericht der interdepertemalen Arbeitsgruppe "künstliche intelligenz" an den bundesrat*, Dec. 13, 2019. [Online]. Available: https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2019/12/bericht_idag_ki.pdf.download.pdf/bericht_idag_ki_d.pdf (visited on 08/01/2021).

[126] Schweizerische Eidgenossenschaft, *Strategie digitale schweiz*, Sep. 11, 2020. [Online]. Available: https://www.bakom.admin.ch/dam/bakom/de/dokumente/informationsgesellschaft/strategie/strategie_digitale_schweiz.pdf.download.pdf/Strategie-DS-2020-De.pdf.

[127] CNAI. "Medienmitteilungen," CNAI. (), [Online]. Available: https://cnai.swiss/ressourcen/pdf/ (visited on 04/23/2022).

[128] ——, "Competence network for artificial intelligence - CNAI," Competence Network for Artificial Intelligence - CNAI. (), [Online]. Available: https://cnai.swiss/en/ (visited on 04/23/2022).

[129] ——, "Project database," CNAI. (), [Online]. Available: https://cnai.swiss/en/products/project-database/ (visited on 04/23/2022).

[130] Swiss Digital Initiative. "Digital trust label & more | swiss digital initiative." (), [Online]. Available: https://www.swiss-digital-initiative.org//our-work (visited on 04/03/2022).

[131] ——, "Label criteria," Digital Trust. (), [Online]. Available: https://digitaltrust-label.swiss/criteria/ (visited on 04/03/2022).

[132] SERI, *Guidelines on artificial intelligence for the confederation - general frame of reference on the use of artificial intelligence within the federal administration*, Nov. 2020. (visited on 10/26/2021).

[133] "Kanton Zürich lanciert eine Testumgebung für Vorhaben im Bereich der Künstlichen Intelligenz," Kanton Zürich. (Mar. 15, 2022), [Online]. Available: https://www.zh.ch/de/news-uebersicht/medienmitteilungen/2022/03/kanton-zuerich-lanciert-eine-testumgebung-fuer-vorhaben-im-bereich-der-kuenstlichen-intelligenz.html (visited on 04/03/2022).

[134] SATW (Schweizerische Akademie der Technischen Wissenschaften), *Künstliche intelligenz: In industrie und dienstleistungssektor*, Oct. 2019.

[135] M. R. Hillemann, "Overview: Cause and prevention in biowarfare and bioterrorism," *Vaccine*, vol. 20, pp. 3055–3067, 2002.

[136] S. Riedel, "Biological warfare and bioterrorism: A historical review," *BUMC Proceedings*, vol. 17, pp. 400–406, 2004.

[137] Z. F. Dembek, L. A. Smith, and J. M. Rusnak, "Botulism: Cause, effects, diagnosis, clinical and laboratory identification, and treatment modalities," *Disaster Medicine and Public Health Preparedness*, vol. 1, no. 2, pp. 122–134, Nov. 2007, Publisher: Cambridge University Press, ISSN: 1935-7893, 1938-744X. DOI: 10.1097/DMP.0b013e318158c5fd. [Online]. Available: https://www.cambridge.org/core/journals/disaster-medicine-and-public-health-preparedness/article/botulism-cause-effects-diagnosis-clinical-and-laboratory-identification-and-treatment-modalities/3AD0CFD5CEF61CAE77923B5F1CA7B908 (visited on 11/29/2021).

[138] A. J. McCluskey, A. J. Olive, M. N. Starnbach, and R. J. Collier, "Targeting HER2-positive cancer cells with receptor-redirected anthrax protective antigen," *Molecular Oncology*, vol. 7, no. 3, pp. 440–451, Jun. 1, 2013, ISSN: 1574-7891. DOI: 10.1016/j.molonc.2012.12.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574789112001299 (visited on 11/29/2021).

[139] B. Erickson, R. Singh, and P. Winters, "Synhetic biology: Regulating industry uses of new biotechnolgies," *Science*, vol. 333, pp. 1254–1256, Sep. 2, 2011.

[140] J. Cohen, "What now for human genome editing?" *Science*, vol. 362, pp. 1090–1092, Dec. 7, 2018.

[141] E. S. Lander, F. Baylis, F. Zhang, *et al.*, "Adopt a moratorium on heritable genome editing," *Nature*, vol. 567, no. 7747, pp. 165–168, Mar. 2019, Bandiera_abtest: a Cg_type: Comment Number: 7747 Publisher: Nature Publishing Group Subject_term: Genetics, Medical research, Ethics, Policy. DOI: 10.1038/d41586-019-00726-5. [Online]. Available: https://www.nature.com/articles/d41586-019-00726-5 (visited on 11/29/2021).

[142] S. Shanaev, S. Sharma, B. Ghimire, and A. Shuraeva, "Taming the blockchain beast? regulatory implications for the cryptocurrency market," *Research in International Business and Finance*, vol. 50, 2020.

[143] Schweizerische Eidgenossenschaft, *Rechtliche grundlagen für distributed ledger-technologie und blockchain in der schweiz*, Dec. 14, 2018.

[144] Eidgenössische Kommission gegen Rassismus. "EKR : TANGRAM 38 - Diskriminierung am Arbeitsplatz: Bei wem liegt die Beweislast?" (), [Online]. Available: https://www.ekr.admin.ch/publikationen/d687.html (visited on 05/10/2022).

[145] Sachverständigenbüro Mülot GmbH. "Pflichten des Verantwortlichen und des Auftragbearbeiters," Das Schweizer Datenschutzgesetz. (), [Online]. Available: https://dsg.ch/kapitel03 (visited on 05/10/2022).

[146] Bundesamt für Justiz. "Stärkung des Datenschutzes." (Mar. 3, 2022), [Online]. Available: https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html (visited on 05/10/2022).

[147] Federal Trade commission. "FTC takes action against company formerly known as weight watchers for illegally collecting kids' sensitive health data," Federal Trade Commission. (Mar. 4, 2022), [Online]. Available: http://www.ftc.gov/news-events/news/press-releases/2022/03/ftc-takes-action-against-company-formerly-known-weight-watchers-illegally-collecting-kids-sensitive (visited on 05/10/2022).

[148] Microsoft. "AI fairness checklist," Microsoft Research. (), [Online]. Available: https://www.microsoft.com/en-us/research/project/ai-fairness-checklist/ (visited on 04/30/2022).

[149] PWC, "Responsible AI - maturing from theory to practice," p. 34, 2021. [Online]. Available: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/pwc-responsible-ai-maturing-from-theory-to-practice.pdf.

[150] EY, *Making artificial intelligence and machine learning trustworthy and ethical*, 2021. [Online]. Available: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/consulting/ey-making-artificial-intelligence-and-machine-learning-trustworthy-and-ethical.pdf.

[151] Google. "Responsible AI practices," Google AI. (), [Online]. Available: https://ai.google/responsibilities/responsible-ai-practices/ (visited on 04/30/2022).

Appendix

A Thesis One-Page Overview

Thesis Overview

The Regulatory Situation of AI in Switzerland

AUTHOR

Iris Amrein
CRM Business Engineer & Technical Consultant
linkedin.com/in/irisamrein

SUPERVISOR

Dr. Donnacha Daly
Head of AI & Machine Learning Studies
linkedin.com/in/donnacha

PROJECT

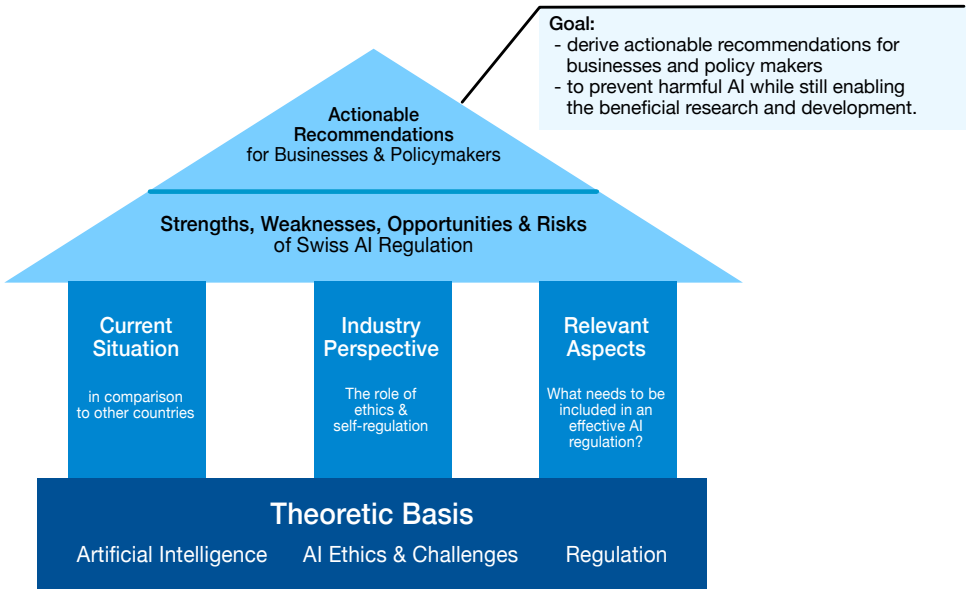
AI finds application in almost every industry today - from farming to banking. But where there are good use cases, there are often also harmful cases. Some companies and governments have started to react to this by drafting regulations. In fact, the EU has drafted an international AI Act that will impose clear restrictions and guidelines on AI applications. On the other hand, the Swiss government has been rather quiet about this topic.

So the questions arise: What should Switzerland do in terms of AI regulation and how does the industry feel about the situation?

TOPICS / STRUCTURE

The thesis first builds a theoretic basis through literature research before exploring the current situation of AI regulation in Switzerland and other countries. A survey and expert interviews will provide information about the industry's perspective on AI regulation as well as which aspects need to be included in a regulatory framework.

The information will be distilled into a SWOT matrix for Swiss AI regulation, which will help derive actionable recommendations for businesses and policy makers.



B Survey Questionnaire

Survey AI Regulation

Section 1/3: Your Involvement With AI

1) **Which Industry do you identify with?** *Choose one.*
   - ☐ Energy
   - ☐ Materials
   - ☐ Industrials
   - ☐ Consumer Discretionary
   - ☐ Consumer Staples
   - ☐ Health Care
   - ☐ Financials
   - ☐ IT
   - ☐ Communication Services
   - ☐ Utilities
   - ☐ Real Estate
   - ☐ Public Sector
   - ☐ Other

2) **What is your position?** *Choose one.*
   - ☐ Executive / Management with technology focus
   - ☐ Executive / Management with business focus
   - ☐ Technology Team Member
   - ☐ Business Team Member
   - ☐ Other (please specify)

3) **Is your company using AI in your Operations, Products and/or Services?** *Choose one.*
   - ☐ Yes
   - ☐ Not Yet, but planning to
   - ☐ No
   - ☐ Never

Section 2/3: Current AI Regulation

4) **Would you consider yourself an expert on AI Regulation in Switzerland or elsewhere?** *Choose one.*
   - ☐ Yes
   - ☐ No

5) **In Switzerland, there is no law to regulate AI technology itself. The government relies on current laws to regulate challenges arising from AI. This is called a technology-neutral approach. Do you think we should have more specific AI regulation in Switzerland?** *Choose one.*
   - ☐ Yes, we should have an AI-regulation & AI-strategy
   - ☐ Yes, we should have a light-weight AI regulation
   - ☐ No, I'm happy with the current approach
   - ☐ No, we should have less regulation around AI

6) **If your organization has a Corporate Social Responsibility (CSR) policy, is AI a part of it?** *Choose one.*

- ☐ Yes
- ☐ No
- ☐ No, but there are other instruments that regulate our use of AI
- ☐ We don't have a CSR
- ☐ I don't know

7) **(Optional) What aspects of AI does your CSR/self-regulation cover?** *Open Text Question*

8) **Have you already been in a situation where you overcame a controversial/ethical issue that arose in the development/implementation of an AI solution?** *Choose one.*

- ☐ Yes
- ☐ No

9) **(Optional) How did you overcome that issue? What did you do to solve it?** *Open Text Question*

10) **Do you personally agree with the following statement: "I believe that my company should be trying to get more AI into the business"?** *Choose one.*

- ☐ Strongly agree
- ☐ Agree
- ☐ No opinion
- ☐ Disagree
- ☐ Strongly disagree

Section 3/3: Possible Future of AI Regulation

11) **Are you in favor of a certifiable standard or label that would identify "responsible AI" products?** *Choose one.*

- ☐ Yes
- ☐ No

12) **(Optional) Why are you (not) in favor of such a standard/label?** *Open Text Question*

13) **Do you believe the implementation of AI regulation at the governmental level or rather self-regulation at the industry level will strengthen the public's trust in AI products & services?** *Choose one.*

- ☐ Yes, at both levels
- ☐ Only at the governmental level
- ☐ Only at the industry level
- ☐ No

14) **Under the EU's regulation, AI applications whose intended purpose may pose "a high risk to the health and safety or fundamental rights of natural persons" need to pass certain requirements to be allowed on the EU market. Do you believe your organization would therefore voluntarily choose to follow the EU's approach although the Swiss approach might be less restrictive?**
Choose one.

- ☐ Yes
- ☐ No

C Survey Answers to open Questions

Answers to open questions in the survey, grouped by topic

WHAT ASPECTS OF AI DOES YOUR CSR/SELF-REGULATION COVER?

Data Privacy, Ethics

- Data privacy and AI
- Ethics, data governance
- Ethics, Data Law Protection
- No use of personal data without consent. No use of facial recognition technologies.
- No discrimination. Transparency in the data and decision-making process. We use only technology we can explain. Transparently report use of AI.
- Business (AI) ethics

How to work with AI

- We are operationalizing responsible AI across our company through a central effort and urge our employees to apply responsible AI principles in their day-to-day work.
- Our other regulations cover how to communicate our own opinions on AI and its usage.

HOW DID YOU OVERCOME AN ETHICAL ISSUE WITH AI? WHAT DID YOU DO TO SOLVE IT?

Abandoning the project

- We did not work with that specific partner. We thought it would not align with our believes and declined the job.
- Refused to participate in customer project
- We removed the feature.

Ensure legality

- help from legal expert
- Users need to give their agreement

Improve the algorithm

- Integrate measures and tests into model evaluation
- Two things: 1. analysis of this specific problem and resolving it with a classical SE approach. 2. transforming the solution into a generic rule - as far as possible.

Build more clarity

- We had to explain the terminology what is meant by "AI" ("marketing term") vs what our technology really was ("efficient automation").

Prevent issues

- We believe in data privacy. Therefore, we leverage privacy preserving machine learning techniques.
- We used our AI Ethics framework to identify and assess the risks to reach our decision to implement or not.
- Being aware of biases in data of AI models (race, gender)

WHY ARE YOU (NOT) IN FAVOR OF A RESPONSIBLE AI STANDARD/LABEL?

In favor

Raises Awareness & Demonstrates Action has been taken

- It demonstrates that the provider of the AI products is aware of and has acted upon possible ethical issues
- To raise awareness of dangers of AI & mitigate them by design

Is needed, because AI can be harmful

- AI usage can sometimes lead, even if not intending to, bias and challenges to privacy, equality.
- Ich denke es wäre auf jeden Fall gut, denn AI kann grosse Auswirkungen haben und ich weiss nicht, ob alle Nutzenden wirklich sehen, dass es auch Gefahren gibt. Allerdings denke ich, dass die Erstellung eines Labels den Fortschritt einschränken wird und die Frage ist auch, wie "grossflächig" wird das Label erstellt. Es müsste wie die ISO-Standard weitreichend sein. Ansonsten ist es sowieso nutzlos.

Can build trust

- There is pros and cons to it. If it is done correctly, it will enable sustainable development. If it is done wrongly, the development will be hindered.
- Having an appropriate evaluation framework to ask the questions in a systematic way is good, for the trust with the non-data literate colleagues and end users.

Others

- Such a label is key to allow for ethical usage of AI. The problem is more: who would issue and certify such a label?
- I am absolutely in favor of responsible AI, but I think that this starts first and foremost with responsible handling of data (data collection, data processing, data transfer).
- only on international standards
- Yes, as long as assessments and results are made public. Otherwise, there is a high risk of white washing as responsible AI is a widely (mis-)used and misunderstood term.

Not in favor

There's no coherent understanding of (responsible) AI

- 1. There is no AI these days, it is impressive statistics or automation, nothing more.
  2. People don't understand "biological" how should they understand "AI"
  3. There is no agreement what AI means it would just cause confusion or cementing some marketing jargon mainly led by big tech companies.
- I doubt that the we have a universal understanding of what responsible AI is.

Slows down innovation

- Such standards do not keep up with development and prevent innovation
- Rules and regulations stifle and slow down innovation

Expensive to maintain

- Such label are expensive to set up and to efficiently maintain, good regulation should do the job

Is a Marketing Tool

- We do not believe in such standards or labels. Those things are often marketing oriented and are not offering real-value to customers. If you "need" a label or a standard, you are not incentivized to do more than needed to get the label - extra efforts are not interesting. Only if you have an intrinsic motivation which is directly related to competitive advantages, you will do whatever is needed.
- The label itself is a form to hide behind what you do and what you have to explain. Without label every piece of software or company should be in full responsibility of what they do.
- Labels are based on self-regulation and often have the tendency to multiply. This is adding complexity and usually does not add much value. Strong internally agreed standards should be the way forward.

AI technology is too vast for a label to encompass it all

- The huge number of possible ways to irresponsibly use AI does not seem likely to be covered by a standard/label.
- Ethical standards should not focus on a specific technology
- I don't think it's the AI itself which is problematic. To try to regulate it might be impossible.

D Interview Guide

AI Regulation in Switzerland

Interview Guide

In this document, you can find the main interview questions, which touch on the subjects I would like to discuss in the interview. It will be a semi-structured interview, allowing flexibility regarding the interview questions.

Context

- The context for the interview is the situation in Switzerland – but feel free to share examples from other countries.
- The term "AI regulation" relates to all measures – binding or non-binding – that attempt to guide businesses to design, develop and deploy AI in a responsible manner.

Interview Questions

Opening / Ice Breaker

1. Can you tell me a bit about yourself and your relation to AI (Regulation)?

Why regulate Artificial Intelligence?

1. Why does AI need to be regulated? (Why not)?
2. Can AI be successful in the future without AI regulation?
3. Do you know any specific examples where a lack or abundance of AI regulation has influenced companies' AI projects (positively or negatively)?

How to regulate Artificial Intelligence?

1. What are the dangers of AI regulation?
2. Are current regulatory approaches suited to regulate AI in Switzerland or other parts of the world?
3. Do you think regulation should come from a governmental or industry level? Why? Is one more efficient than the other?
4. Is there any considerable alternative to AI regulation?

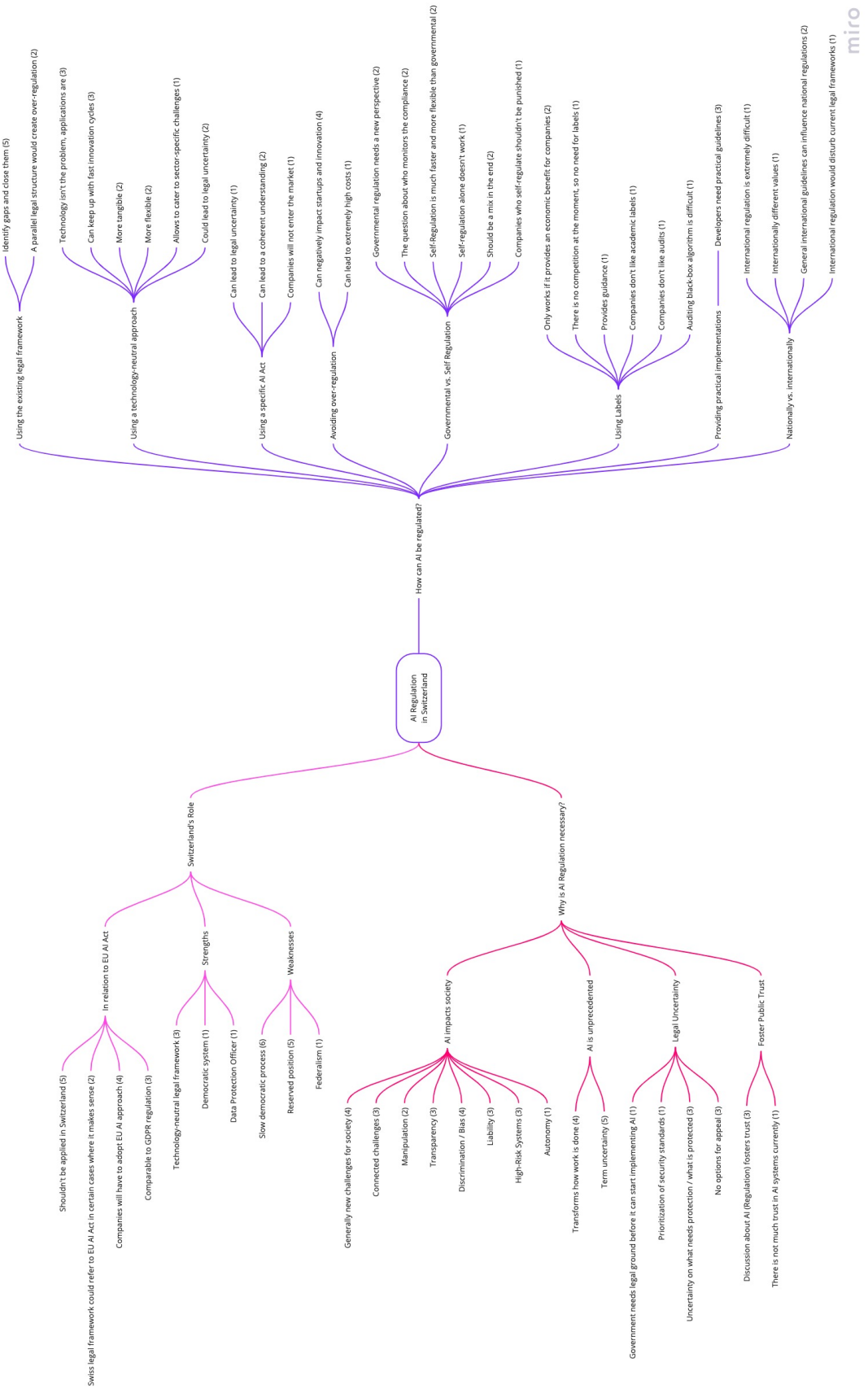
Switzerland and the EU / other countries

1. Do national AI regulations even make sense?
2. If you could make one change to how AI is regulated in Switzerland in a heartbeat and without any repercussions, what would you change?
3. Where do you see Switzerland's strengths and weaknesses in AI regulation?
4. Would it be detrimental to the success of Swiss AI companies if we didn't do anything regarding AI regulation (especially in terms of the EU's AI regulation)?

Outlook / Closer

1. How do you see the situation of AI (regulation) in 20-50 years?

E Interview Analysis - Mind Map


How can AI be regulated?
AI Regulation in Switzerland
Switzerland's Role
In relation to EU AI Act
Shouldn't be applied in Switzerland (5)
Swiss legal framework could refer to EU AI Act in certain cases where it makes sense (2)
Companies will have to adopt EU AI approach (4)
Comparable to GDPR regulation (3)
Strengths
Technology-neutral legal framework (3)
Democratic system (1)
Data Protection Officer (1)
Weaknesses
Slow democratic process (6)
Reserved position (5)
Federalism (1)
Why is AI Regulation necessary?
AI impacts society
Generally new challenges for society (4)
Connected challenges (3)
Manipulation (2)
Transparency (3)
Discrimination / Bias (4)
Liability (3)
High-Risk Systems (3)
Autonomy (1)
AI is unprecedented
Transforms how work is done (4)
Term uncertainty (5)
Legal Uncertainty
Government needs legal ground before it can start implementing AI (1)
Prioritization of security standards (1)
Uncertainty on what needs protection / what is protected (3)
No options for appeal (3)
Foster Public Trust
Discussion about AI (Regulation) fosters trust (3)
There is not much trust in AI systems currently (1)
Using the existing legal framework
Identify gaps and close them (5)
A parallel legal structure would create over-regulation (2)
Using a technology-neutral approach
Technology isn't the problem, applications are (3)
Can keep up with fast innovation cycles (3)
More tangible (2)
More flexible (2)
Allows to cater to sector-specific challenges (1)
Could lead to legal uncertainty (2)
Using a specific AI Act
Can lead to legal uncertainty (1)
Can lead to a coherent understanding (2)
Companies will not enter the market (1)
Avoiding over-regulation
Can negatively impact startups and innovation (4)
Can lead to extremely high costs (1)
Governmental vs. Self Regulation
Governmental regulation needs a new perspective (2)
The question about who monitors the compliance (2)
Self-Regulation is much faster and more flexible than governmental (2)
Self-regulation alone doesn't work (1)
Should be a mix in the end (2)
Companies who self-regulate shouldn't be punished (1)
Using Labels
Only works if it provides an economic benefit for companies (2)
There is no competition at the moment, so no need for labels (1)
Provides guidance (1)
Companies don't like academic labels (1)
Companies don't like audits (1)
Providing practical implementations
Auditing black-box algorithm is difficult (1)
Developers need practical guidelines (3)
Nationally vs. internationally
International regulation is extremely difficult (1)
Internationally different values (1)
General international guidelines can influence national regulations (2)
International regulation would disturb current legal frameworks (1)
miro

F Interview Analysis Overview

Overview Interviews

Topic	Category	Code	Sum
Why is AI Regulation necessary?	AI impacts society	General new challenges for society	4
		Transparency	3
		Discrimination / Bias	4
		Liability	3
		High-Risk Systems	3
		AI challenges are equally important	3
		Manipulation	2
		Autonomy	1
	AI is unprecedented	Term Uncertainty	5
		Transforms how work is done	5
	Legal Uncertainty	No options to appeal	4
		People and companies don't know what to protect	3
		Prioritize security standards	1
		Government needs legal ground before implementing AI	1
	Foster Public Trust	Discussion about AI Regulation fosters trust	4
		Not much trust in AI systems	1
How can AI be regulated?	Use existing legal framework	Identify gaps and close the gaps	6
		A parallel new legal structure would create confusion	2
	Using a technology neutral approach	AI technology isn't the problem – applications are	3
		More suited for fast innovation	3
		More flexible	2
		More tangible	2
		Could lead to legal uncertainty	2
		Allows to cater to sector-specific challenges	1
	Using a specific AI Act	Can lead to coherent understanding	2
		Companies will not enter the market	1
		Can lead to legal uncertainty	1
	Not too strict	Over-regulation negatively impacts innovation / startups	4
		Can lead to extremely high costs	1
	Governmental vs. self-regulation	Governmental regulation needs new perspective	2
		Self-regulation is much faster	2
		Should be a mix of governmental and self-regulation	2
		Question is who monitors compliance	1
		Self-regulation hasn't worked	1
		AI has no priority	1

How can AI be regulated?	Governmental vs. self-regulation	Self-regulating companies are "punished"	1
	Using Labels	Only works if benefit is provided	2
		There is no need for comparison yet	1
		Provides guidance and is a good starting point	1
		Companies don't like academic labels	1
		Companies don't like audits	1
		Auditing will be difficult	1
	Providing practical implementation	Developers need practical guidelines	2
		Practical questions or guidelines for every step in the process	1
	Nationally vs. internationally	International regulation is difficult	2
		General international guidelines can influence national regulations	2
		Different values	1
		International regulation would disturb current regulation	1
What is Switzerland's role in AI regulation?	Weaknesses	Slow democratic process	6
		Reserved position	5
		Federalism	1
	Strength	Technology-neutral framework	3
		Democratic system	1
		Data Protection Officer	1
	In relation to EU AI Regulation	Shouldn't be applied in Switzerland	5
		Companies will have to adopt AI approach	4
		Comparable to GDPR regulation	3
		Swiss legal framework could refer to EU regulation	2

Codes and their respective statements

WHY IS AI REGUALTION NECESSARY?

AI Impacts Society

General new challenges for society

- AI poses many challenges
- AI has many implications for society
- Implications for society have to be anticipated and negative implications should be averted
- AI can have implications on fundamental rights

Manipulation

- Can be very high-risk because such systems can hurt people and have negative consequences like in election s and voting
- We need to regulate AI to protect democracy – manipulation can be very polarizing and hurt democratic discourse

Transparency

- Transparency is needed from a governance perspective because most legal laws require transparent governmental actions. Explainability & traceability are a challenge
- If a system is more transparent, we can fight discrimination better because we can identify which input resulted in a discriminating output
- A person should know that and how AI is applied if they're interacting with/affected by it

Discrimination/Bias

- Outputs can be discriminating, this has to be countered
- I think regulation is certainly necessary, especially when it comes to discrimination
- Unfair bias and discrimination is one of the most pressing issues and where most would see a bigger problem – it can lead to massive reputational losses
- Almost all our projects were in bias/discrimination

Liability

- Product liability is very focused on tangible products, not on software
- Product liability still focuses on tangible products – far form data or software. There's differentiation between material and immaterial goods. In Switzerland and the EU, there is consensus that software falls under the product liability law
- AI is not just software code

High-risk-systems

- When talking about AI Regulation, we often exclude autonomous weapons. If I could, I would ban them internationally
- I would ban high-risk applications like facial recognition and autonomous weapons – we can agree on some things we don't want AI to be able to do
- High-risk systems and general purpose systems which can cause massive damage should be regulated first

All challenges are equally important

- We cannot look at the challenges isolated or should prioritize one above the other
- There are so many applications domains where AI can be implemented. Legislature cannot possibly grasp every single interaction possibility with AI and regulate it
- There is no challenge that is generally the most important. It's sector-/application domain-specific

Autonomy

- The more autonomous the more difficult it is to create accountability

AI is unprecedented

Term uncertainty

- AI is a very broad term. It includes many different technologies and is continuously developing
- We first need to identify all challenges or problems that AI poses
- We have to understand what AI is technically and what it can and will be able to do. There is a huge gap between what is possible today and what will be possible in the future.
- It's already a challenge to understand from a legal perspective what AI is and what law can even do effectively in this regard
- AI is very abstract and there are different understandings. AI can not be generalized so it shouldn't be regulated as one
- Defining AI for regulation is very important. Otherwise it can lead to many dangers.

Transforms how work is done

- AI is a core technology for our business
- It's a new kind of technology
- Suddenly there are not only humans doing the work anymore but machines as well
- The speed and scale of what AI can do is unprecedented and that warrants regulation
- Humans don't deal well with things that can evolve exponentially

Legal Uncertainty

No options to appeal wrong algorithmic decisions

- There are not many options to act against AI usage. If you know about an AI application you don't want to interact with, you can only ignore it
- There need to be possibilities for affected people to defend themselves
- If I could, I would immediately create options for the affected to get an explanation about the decision and the ability to appeal such decisions
- People interacting with AI should have the possibility to get help. False decisions should be able to be checked and appealed

Prioritize security standards

- In sensitive domains like health care or self-driving cars we need to prioritize security standards from the beginning

Government needs legal grounds to use AI

- The government mustn't apply AI as long as there are no legal grounds for it. So far there are no legal grounds that allow something like AI or algorithmic-based systems – even if there is a human making the decision in the end.

People and Companies don't know needs to be protected

- It is important for companies to have some degree of certainty – or at least clear rules.
- Could be that regulation helps with more legal certainty for startups so they can make better decisions
- Companies cannot understand fragmented legal framework. They want guidelines / checklist to know exactly what is important in regards to AI in their domain. The same goes for the public. They are wondering whether and how they are protected

Foster Public Trust

Discussion about AI Regulation fosters trust

- How exactly the law is drafted does not impact trust, more the discussion about AI influences trust
- The discussion about there being AI problems and that AI is implemented already foster trust in such systems and that they see that they have the options to act against it
- Regulation can also build trust among consumers
- the main purpose of regulation is to mitigate harm but it can also help foster trust in the technology
- A regulation can start discussions and foster trust and it can help include stakeholders
- Society needs to know that they can draw a line and they can say when they don't want something automated

There is not much trust in AI

- Many people don't trust AI systems by default. If we regulate, we can foster trust in such systems

HOW CAN AI BE REGULATED

Using an existing approach

Identify Gaps and close them

- Work with norms we already have
- Many experts say that we don't need an AI Act but we should work with norms we already have. Maybe we need to interpret them differently or work with consensus from academia
- there are already many regulations i.e. for vehicles, medicinal products. We have to see if there are gaps that could be closed with special regulations in those sectors. There are few gaps that need to be closed, says EDA
- There are already many laws that protect people from challenges that AI poses.
- A lot of aspects are already incorporated in the current legal framework. It is sector specific for the different industries and application domains
- There are many laws that could be adapted or improved for AI applications

A parallel new legal structure would create confusion

- I would start with the current regulation and close gaps where needed to prevent over-regulation
- AI applications are embedded into a current system so they should be regulated in the same way the current system is now - and not in a parallel legal structure

Using a technology neutral approach

AI technology isn't the problem – application of it is

- Problems that arise with AI technology are technology-independent problems -> i.e. with Facial Recognition it's not the problem that the face is identified but that society is monitored.
- There are so many application domains where AI can be implemented. Legislature cannot possibly grasp every single interaction possibility with AI and regulate it
- We shouldn't regulate technology but the outcomes. The focus is currently too much on the technology and measurability than on application

More suited for fast innovation cycles

- It's possible that when AI regulations go into effect there is already a new technology that can do other things and has its own problems
- the legal process can't hold up with the quick innovation cycles
- The Swiss legislative process is too slow for the quick digital technology development cycles

More flexible

- That something is described in a technology-neutral way is always because it needs to be flexible
- Is very open and offers possibility to react more neutral

More tangible

- I think it's more tangible for most people. The public does not really know this technology.
- We should focus on end-user of such systems and not on technology or products themselves

Allows to cater to sector-specific challenges

- The regulatory / ethical approach is different in each sector (health care vs. Vehicles vs. Government)
- Is better than regulating a group of possible application domains

Could lead to legal uncertainty

- Companies don't know where to look
- A risk with the technology-neutral approach is that it could become very fragmented. We could lose oversight to find all the relevant elements – especially given the exponential development of AI technology

Using specific AI Act

Can lead to coherent understanding

- Leads to better understanding of the rules for businesses - more practical
  - For example, with Social Media the legal framework already covered most of the use cases. But companies wanted something more explicit. Also for communication purposes to show that they are aware of the topic, they have regulated it and that they know what the "rules" are
- All market players have to follow the same rules so self regulating businesses are not punished

Companies will not enter the market

- It's possible that Swiss businesses will just not provide certain services that would fall under the EU Act in the EU market. The question is "is it lucrative to treat EU differently?"

Can lead to legal uncertainty

- A new AI act can lead to legal insecurity because people don't know where they have to look for what is applicable in their case
- The EU AI Act is extremely complicated, not many people really understand it

Not too strict

Over-Regulation negatively impacts innovation/startups

- There should always be room for innovation. We want to use these systems as they can be very beneficial. The benefit has to be compatible for society - and that is actually the goal of state regulation, in private law and in the public sector
- Startup landscape could be negatively impacted if the matter is over-regulated
- If AI is defined loosely, many applications could fall in the scope of the law, which means they need to fulfill strict requirements before going to market. This leads to Startups maybe not being able to afford that, so they will go to USA / Asia
- A common argument is the slow-down of innovation and there’s truth in that
- If It's very strict and can be detrimental for startups

Can lead to extremely high costs

- Costs could explode for example for startups → EU's "high-risk applications" have many requirements which cost a lot of money
- Also if liability questions are not answered uniform then companies could expect high costs

Governmental vs. Self-regulation

Governmental regulation needs new perspective

- AI needs to be regarded as a life cycle - so regulation needs to be continuously developed as well
- AI applications are embedded into current legal framework (DSG, discrimination, security) -> constraints the scope for self-regulation already
- AI has a life-cycle - responsibilities over this life cycle need to be clear between adopters and providers

The question is about who monitors compliance

- The question is who monitors if the rules have been followed -> this could be sector-specific
- Regulation is good, but the infrastructure must be ready too (i.e. judges need to know what AI is about)

Self-regulation is much faster

- Self-regulation is much faster than governmental regulation
- Businesses start self-regulation because government is not there yet
- Self-regulation is quicker and more flexible

In the end regulation should consist of governmental and self-regulation

- At the moment it is probably suited more - but in the end it should be a mix of governmental and self-regulation
- The governmental regulation can serve to level the playing field and the self-regulation can then be the rulebook on how the legislation is implemented practically

Self-regulation hasn't worked

- So far, self-regulation hasn't worked

Companies who self-regulate are punished

- there is a need for a "level playing field" - so self-regulating businesses are not punished

AI is not prioritized

- Legislators are too busy with digitalization and data protection, AI is not prioritized

Using Labels

Only works if it provides benefits

- Businesses only adopt labels if there is demand for it or if it has to differentiate from the competition quality wise
- It can add value to a company

There is no need for comparison yet

- It's quite early because there aren't really that many companies in Switzerland that apply AI (with regards to humans, not industrial). But I could see that a label could be interesting for the HR sector

Provides guidance and is a good starting point

- A label is a good starting point and tells companies which questions they must ask when they develop AI
- Labels foster the discussion about the topic
- Can be a kind of guideline to show what needs to be taken into account

Companies don't like academic labels

- Often labels are "too academic" for businesses and too far from real business. So having business people as experts can foster trust of companies in such a label

Companies don't like audits

- Not all companies like labels because they often come with an audit. At an audit gaps could be discovered which then means they need to be closed too

Auditing will be difficult

- We need to be careful about what a label could be – because auditing an algorithm that is more or less a black box is not going to be easy

Providing practical implementations

Developers need practical guidelines

- We've had good feedback to clear ethical AI principles and practices

- The developers identify the topic as important but don't know how and what technical implications from regulation are

Practical questions or guidelines needed in every step of the process

- Regulation should be approached with a process logic: In which step of development need to be asked which questions or what needs to be taken into account?

Nationally vs. Internationally

International regulation is difficult

- It's extremely difficult to regulate something on an international level -> see covid pandemic: there are no uniform health- or pandemic-laws even if a virus doesn't stop at the border
- I think it's not possible to regulate internationally on a political level
- international law is not always followed by everyone. Difficult to know how to react then, i.e. with autonomous weapons

Different values

- There are different legal systems and understandings of democracy and different values that are deemed worth protecting. For example, in China, mass surveillance is the norm, in Switzerland that would never be possible

General international guidelines can influence national regulations

- Governments can orientate on international guidelines like Europarat or OECD, which can lead to some kind of coherence
- It makes sense to have something of the European level and the implementation is then up to the member countries. Local implementation of broad principles that are shared by everyone

International regulation could disturb legal frameworks

- if we draft a "horizontal" liablity law we would risk disturbing the current legal systems

WHAT IS SWITZERLAND'S ROLE IN AI REGULATION?

Weaknesses

Slow Democratic Process

- Legislation is a long process. It often takes years. A general weakness is the democratic process. It takes an incredibly long time.
- The Swiss speed might at some point not be fast enough.
- The Swiss legislative process is very slow (Drafting, consulting, parliamentary vote, popular vote)
- It takes a long time from drafting a new law to it going into effect. That's a problem with the rapid development cycles of digital technologies
- Slow legislative and implementation process
- It's a slow process

Reserved Position

- In principle, Switzerland is rather cautious when it comes to regulation. They prefer to observe what the EU is doing in order to not block their access to the market
- We are more on a reactive approach, following the lead
- Switzerland is still at the beginning, the EU is much further. We're still trying to understand and find use cases
- Culturally Switzerland prefers less regulation to more regulation (like GDPR)
- Switzerland is not as active as other countries like the EU - it just watches what other countries are doing

Federalism

- "Blame Game" between Cantons and State

Strengths

Technology-neutral framework

- Using existing norms could be more comprehensible
- In the long run, not regulating a specific technology but including other technological developments is more sustainable
- We have all the elements we need to regulate all the important aspects of AI
- There are already laws that are also applicable to AI applications. We're not in a completely lawless state

Democratic system

- Democracy is a strength for the government. It is some kind of two-sided sword

Data Protection Officer

- It's a good instrument / instance
- data and ai are quite related so it could be a possibility to include them or such an instrument in AI regulation

In relation to the EU AI Act

Shouldn’t be applied in Switzerland

- In Switzerland such a general AI law is generally rejected
- Most aspects of the EU AI Act are covered by Swiss laws already
- Could negatively affect startup scene if too strict
- EU AI Act is very different to how Switzerland has dealt with technologies in the past. In Switzerland there is a completely different legal framework and organization form so the question is how much should we effectively carry over and what not?
- We don't need an all-new Swiss AI Act

Swiss legal framework could refer to EU AI Act

- Switzerland would need to also instantiate the corresponding structures / administrations
- A risk based regulation with certain requirements before the market entry is a good approach as long as it's not too strict

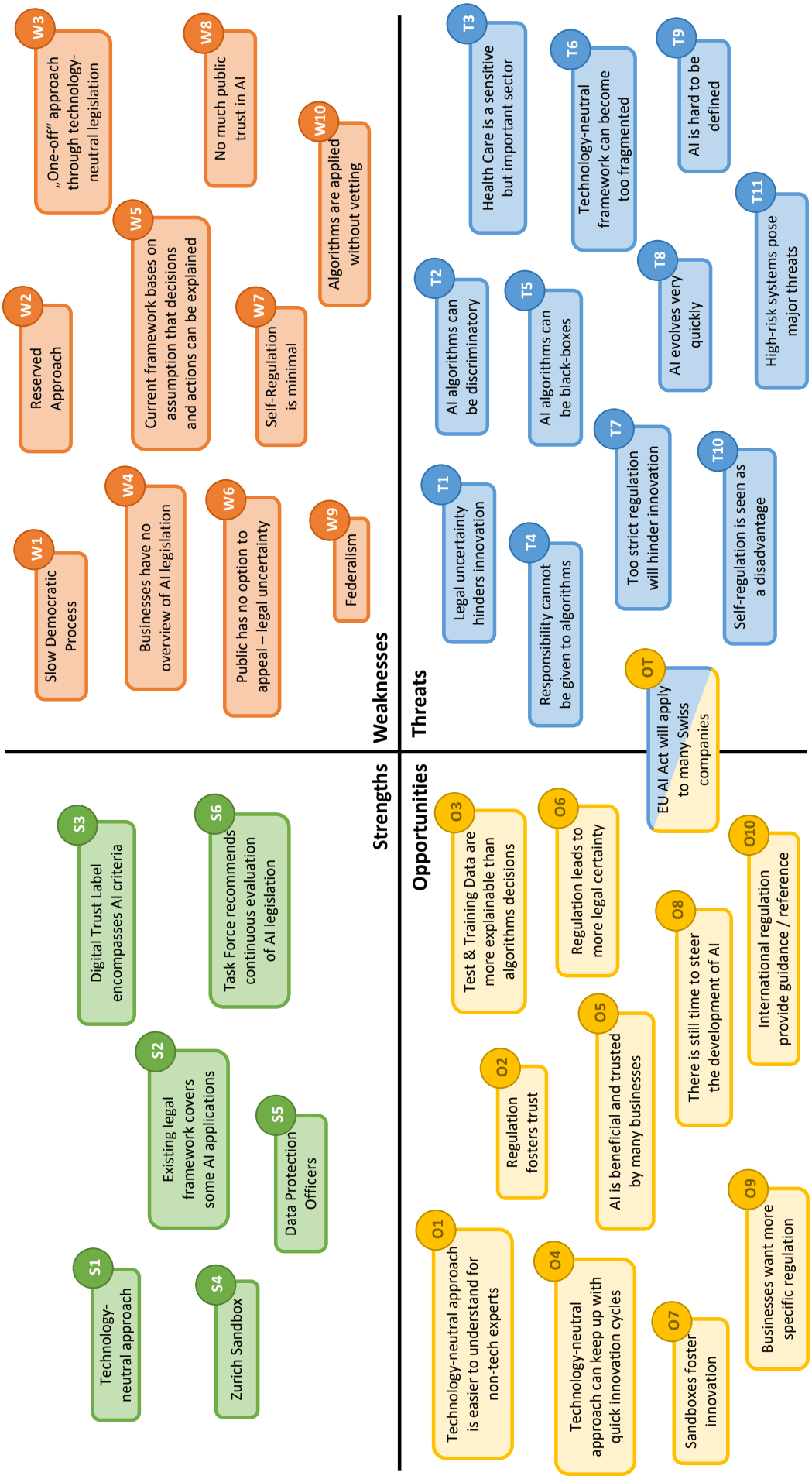
Companies will have to adopt EU AI Approach

- It's possible that Swiss businesses will cater to EU approach to have access to the market
- Swiss Companies selling to the EU will have to comply with the AI Act anyway
- Businesses almost never develop systems just for CH market, usually also EU
- Most companies will have to work with the EU AI Act. Swiss justices will probably also look towards European justices and verdicts

Comparable to GDPR regulation

- EU has defined a certain standard for data protection with GDPR. To enter the EU market, some internal processes had to be adapted generally
- Most businesses are GDPR conform even if DSG hasn't fully been adapted in CH yet
- The Data Protection Law that is coming into effect in Switzerland is very similar to the GDPR guidelines

G SWOT of Swiss AI Regulation


Weaknesses
W1 Slow Democratic Process
W2 Reserved Approach
W3 „One-off" approach through technology-neutral legislation
W4 Businesses have no overview of AI legislation
W5 Current framework bases on assumption that decisions and actions can be explained
W6 Public has no option to appeal – legal uncertainty
W7 Self-Regulation is minimal
W8 No much public trust in AI
W9 Federalism
W10 Algorithms are applied without vetting
Threats
T1 Legal uncertainty hinders innovation
T2 AI algorithms can be discriminatory
T3 Health Care is a sensitive but important sector
T4 Responsibility cannot be given to algorithms
T5 AI algorithms can be black-boxes
T6 Technology-neutral framework can become too fragmented
T7 Too strict regulation will hinder innovation
T8 AI evolves very quickly
T9 AI is hard to be defined
T10 Self-regulation is seen as a disadvantage
T11 High-risk systems pose major threats
OT EU AI Act will apply to many Swiss companies
Strengths
S1 Technology-neutral approach
S2 Existing legal framework covers some AI applications
S3 Digital Trust Label encompasses AI criteria
S4 Zurich Sandbox
S5 Data Protection Officers
S6 Task Force recommends continuous evaluation of AI legislation
Opportunities
O1 Technology-neutral approach is easier to understand for non-tech experts
O2 Regulation fosters trust
O3 Test & Training Data are more explainable than algorithms decisions
O4 Technology-neutral approach can keep up with quick innovation cycles
O5 AI is beneficial and trusted by many businesses
O6 Regulation leads to more legal certainty
O7 Sandboxes foster innovation
O8 There is still time to steer the development of AI
O9 Businesses want more specific regulation
O10 International regulation provide guidance / reference