Unintended Outcomes of Artificial Intelligence in Online Advertising

Lotte I. van Dijk (s2837544)

Faculty of Behavioural, Management, and Social Sciences

Department of Business Administration, University of Twente

MSc BA Thesis

Strategic Marketing and Servitisation

First supervisor: Dr. A. Leszkiewicz

Second supervisor: Dr. Matthias de Visser

Third supervisor: Robert Marinescu-Muster

12th of April, 2023

ABSTRACT

Advertisers increasingly employ Artificial Intelligence (AI) applications within online advertising to create and monitor ads, formulate strategic marketing decisions based on a large amount of customer data and accurately target their customers. It saves time and money and creates a competitive advantage. However, the application of AI within online advertising can sometimes lead to unintended outcomes. Therefore, the European Commission (EC) has issued the Artificial Intelligence Regulation Proposal, known as 'the AI Act', to mitigate harmful AI. This thesis seeks to comprehend the company's viewpoint on employing AI and facing new laws and regulations. Furthermore, it describes mitigation methods to prevent unintended outcomes and includes a review of adopted and proposed laws and regulations. Finally, a checklist is compiled for companies applying AI within their online advertising campaigns to prevent them from creating harmful unintended outcomes. This thesis achieved these contributions by conducting two studies: 1) a case study on Facebook as an AI-driven advertising platform and 2) in-depth interviews about the application of AI in online advertising, unintended outcomes and viewpoint on the AI Act, including 18 participants. The findings result in an overview of how companies employ AI in online advertising. Furthermore, it identifies the advantages, disadvantages, opportunities and threats of applying AI in online advertising from a company's viewpoint. It identifies the potential causes of unintended outcomes and describes mitigation methods to prevent these outcomes from happening. This thesis combines its findings into a checklist for companies applying AI in advertising without creating harmful unintended outcomes.

Keywords: AI Act, Algorithms, Artificial Intelligence, Bias, Discrimination, Law, Legal, Machine Learning, Marketing, Online Advertising, Regulations, Social Media, Unintended Outcomes

2837544 – Lotte van Dijk – Master Thesis – Unintended Outcomes of Artificial Intelligence in Online Advertising

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Ads	Advertisements
AI	Artificial Intelligence
CRISP-DM	Cross Industry Standard Process for Data Mining.
EC	European Commission
EU	European Union
GDPR	General Data Protection Regulation
FRAIA	Fundamental Rights and Algorithm Impact Assessment
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
SMEs	Small Medium Enterprises
STEM	Science, Technology, Engineering, Mathematics

1. INTRODUCTION

Artificial Intelligence (AI) is the capability of a computer application to imitate human reasoning – and behaviour by interpreting and learning from data. It is one of the fastest-emerging technologies, and human society increasingly depends on it (Akter et al., 2022; Ngai & Wu, 2022). Benefits of AI have been experienced in many aspects of life as it has brought revolutionary results in improving labour efficiency, reducing labour costs, optimising the structure of human resources, and creating new job demands (Akter et al., 2022; Ngai & Wu, 2022; Zhang et al., 2022).

AI is being employed in new ways across a variety of industries. In the case of marketing, the application of AI presents promising results because of expanding computing power, the availability of big data, and the development of machine learning algorithms and models. There are wide applications of AI in marketing. Amazon employs AI to automate its shipping and delivery, Starbucks uses AI to discover and engage new customers, and Walmart applies Natural Language Processing (NLP) to train its chatbot to interact with clients online (Huang & Rust, 2021; Shah et al., 2020).

In particular, introducing AI within online advertising has led to enormous financial success (Ali et al., 2019). The application of AI within online advertising allows companies to introduce (potential) customers to their products and/or services to stimulate the purchase of their products and/or the employment of their services. Moreover, companies employ AI to auction keywords, target customers and create content. Applying AI within online advertising increases efficiency, improves conversion rates, and leads to better customer insights, understanding, and satisfaction (Shah et al., 2020). Digital ad platforms such as Google (Alphabet) and Facebook (Meta) employ AI to deliver these ads to a specific audience based on the advertiser's chosen specifications. Customers interact with advertisements that appeal to them, and companies can affordably market their products/services (Chhillar & Aguilera, 2022; Shah et al., 2020).

However, studies have proven that ad delivery can lead to unintended outcomes. Applying AI within ad delivery activities can limit users' interaction with specific ads based on gender or race due to biased algorithmic optimisation (Ali et al., 2019; Bechmann & Bowker, 2019; Cabañas et al., 2018; Datta et al., 2018; Imana et al., 2021). Among other studies, Ali et al. (2019) and Asplund et al. (2020) found that by running similar employment ads (different

jobs, customised text, but the same bidding strategy and same target audience), Facebook delivered the ads to its audiences based on biased gender and racial lines.

Biased AI leads to discrimination that escapes the current Discrimination – and Data regulation laws (European Digital Rights, 2021). Therefore, the EC has proposed the Artificial Intelligence Regulation Proposal (AI Act). The first European law on AI is meant to be a legal framework to prevent harmful unintended outcomes caused by AI (Bringas Colmenarejo et al., 2022; European Commission, 2021). However, the literature suggests that the AI Act contains several loopholes and exceptions. In addition, companies applying AI within their business operations struggle with meeting the requirements of the AI Act (Borsci et al., 2022; Bringas Colmenarejo et al., 2022; Veale & Borgesius, 2021).

This thesis seeks to comprehend the company's viewpoint on employing AI and facing new laws and regulations. It accomplishes so by exploring the advantages, disadvantages, opportunities and threats of adopting AI in online advertising from the perspective of these companies. As a result, it tries to identify how companies counter unintended AI outcomes. Finally, this thesis proposes a checklist for companies applying AI in advertising without potentially harmful unintended outcomes. To reach these objectives, the following research questions have been developed:

Table 1

Research Questions (Own elaboration)

Research Question	Empirical Approach		
	Literature review	Case study	Interviews
How are companies implementing Artificial Intelligence within their advertising campaigns?	YES	YES	YES
To what extent does employing Artificial Intelligence lead to unintended outcomes in online advertising?	YES	NO	YES
How are companies employing Artificial Intelligence for their advertising campaigns currently complying with a legal framework(s) in the EU?	YES	NO	YES

This thesis uses a qualitative descriptive methodology to answer these research questions. This thesis conducts a case study on Facebook as an AI-driven advertising platform and depth interviews to answer its research questions. The in-depth interviews are held with 18 AI, marketing, and legal experts.

The contributions of this thesis are assessing legal frameworks in the EU based on their ability to counter unintended biased outcomes of the usage of AI in online advertising campaigns; Implementing the companies' perspective on the new AI Act by conducting case studies and in-depth interviews; Outlining the shortcomings of existing laws and regulations in countering unintended outcomes of applying AI by companies, and finally developing a checklist that companies can consult to prevent unintended outcomes from happening within their online advertising activities.

The structure of this thesis is as follows. First, the theoretical background is outlined. Following that, the two types of qualitative research methodologies and their results will be discussed in Chapter 2 and Chapter 3. These results will be summarised in Chapter 4, which will present the aforementioned checklist by summarising and integrating the results of the two studies. Next, in Chapter 5, the results and their limitations will be highlighted, and potential future research topics will be suggested. Lastly, in Chapter 6, the conclusions and recommendations will be drawn.

To provide an answer to the research questions, this Chapter continues by describing how AI can be applied in online marketing. Then, after defining AI and how it can result in unintended outcomes, this Chapter gives an overview of the laws and regulations the European Commission (EC) has proposed or implemented. This Chapter's final section focuses on how these laws and regulations can help to prevent biases.

1.1.   How is Artificial Intelligence applied in Online Marketing?

The employment of AI in business practices is on the rise. The increasing availability of big data and – computer power, combined with the decreasing computing costs, led to expanding importance of AI (Huang & Rust, 2021; Ngai & Wu, 2022). Even more, it has been claimed that AI will significantly affect marketing in the future (Davenport et al., 2020). Marketers increasingly depend on AI for communication, delivering value and building a connection with their customers.

AI focuses on personalisation and creating value for companies and customers. Companies concentrate more on understanding customers' preferences to determine their marketing mix (Gupta et al., 2020). Among other things, AI is being employed for data analysis: these tools can automate analytical models that can be employed for the data mining large sets of (consumer) data. This provides marketers with new insights and opportunities, which are then applied to improve the performance of marketing activities and operations (Akter et al., 2022; Ngai & Wu, 2022). Table 2 elaborates on how AI has developed online advertisement within the 4Ps marketing mix. Because this thesis focuses on unintended outcomes experienced by people, the P for People has also been added to this table (Bahador, 2019; Huang & Rust, 2021; Ngai & Wu, 2022).

The benefits of AI seem limitless. However, marketing literature suggested that through biased datasets that represent the discriminating judgements of society or through biases in the training of AI programs, AI could lead to discriminating outcomes (Akter et al., 2022; Chhillar & Aguilera, 2022; van Giffen et al., 2022; Zuiderveen Borgesius, 2018). Discrimination caused by AI has also been detected in online marketing (Ali et al., 2019). Borgesius (2018) explored how AI could lead to discrimination and highlighted how people could be discriminated against without awareness, especially in advertising. If an AI system targets male users in a job advertisement, women do not realise that they, as female users, are being excluded from this ad campaign. In short, a Facebook user would not know about the existence of an ad if they do not interact with specific ads. Similarly, Asplund et al. (2020) and Lambrecht & Tucker (2019) have proven that advertisements were delivered based on biased gender and racial lines. However, Ali et al. (2019) and Dalenberg (2018) explored how ads may be delivered in a way that has led to unintended outcomes in online advertising. This study found that a marketeer does not intend to exclude particular users from interacting with their ad. The causes and types of biases will be further discussed in Chapter 1.3.

Table 2

4P's Marketing Mix (Own elaboration)

4Ps Marketing Mix			
Product	**Market Analysis** What products is the customer buying? What products are the competitors offering? What products are now being well sold within the market?	**Purchase Decision Prediction** Predict trends (opportunities and threats) within the market.	**Product Recommendation** Offer personalised trends for customers.
Price	**Automatic payment** Payment through online payment software, such as Apple Pay, Google Pay, PayPal, etc. (Huang & Rust, 2018; Ngai & Wu, 2022).	**Personalised pricing** The employment of personalised algorithms to implement price discrimination based on a customer's buying behaviour. (Yu et al., 2020)	**Price negotiation** By learning algorithms and negotiation techniques, AI systems can now negotiate the price of products and services with customers. (Yadav et al., n.d.)
Promotion	**Advertising Management** Keeping track of which advertisement is performing well and which one is not.	**Demand prediction** Predict the demand and sales of customers in marketing activities and adjust the distribution of marketing resources to keep the effectiveness of marketing activities.	**Chatbots** Communicate at a human level with customers through the help of a text-based dialogue system.
Place	**Self-check-in and check-out** Members can employ their membership cards to check in when entering a plane or gym (Martínez et al., 2020).	**Personalised frontline interactions** Using a Personalised Shopping App allows shoppers to ask questions regarding products within this store (Samuely, 2017).	**Personalised experience in customer engagements** Stores employ facial recognition to recognise customers (Low & Clifford Lee, 2021).

| **People** | **Churn Prediction** Find out which customers will most likely cancel /her subscription. (Ahmad et al., 2019).

Identifying Customer Segments It identifies patterns and creates customer segments by dividing a market based on demographic, geographic, and sociographic factors. (Huang & Rust, 2021; Ngai & Wu, 2022). | **Target Customer Prediction** Predicting whom your customers are going to be. Companies can decide whom to target by applying predictions based on the characteristics highlighted in the Product Recommendation but also through demographic factors and customer loyalty. (Ahmad et al., 2019). | **Engagement** Customer engagement by applying an algorithm that is based on NLP (Natural Language Processing) that explores social media content and how this affects customer engagement (Dolan et al., 2019; Lee et al., 2018; Tafesse & Wien, 2017) |

1.2.   What is Artificial Intelligence, and what are the types of AI Algorithms?

This thesis primarily focuses on the application of AI in online advertising and how it can lead to unintended outcomes. Because of this, it will briefly discuss the connections between AI, ML, and algorithms without going into great detail regarding their functions. The focus of this thesis will nonetheless remain on AI.

"Artificial intelligence is a distinctive capability of computer applications that can exhibit characteristics of human intelligence through interpreting external data and exhibiting flexible adaptation through learning from data." – (Akter et al., 2022, p.202)

AI can integrate emotional cognition, human-computer interaction, storage of data and making decisions (Zhang & Lu, 2021). AI employs computers to simulate human intelligent behaviours, such as learning, judging and making decisions. To operate, AI uses one or more Machine Learning algorithms. Machine Learning (ML) is a subfield of AI that can identify patterns in big data and is employed for decision-making (Kumar et al., 2020). Therefore, it is applied in different industries, such as marketing, accounting, finance and customer service (Ngai & Wu, 2022). ML methods are built on mathematical models and algorithms, such as neural networks, decision trees, linear regression models etc., and are applied to data and computer applications for predicting, clustering, classification and reduction (Akter et al., 2022; Zhang & Lu, 2021). Table 3 presents examples of algorithms that can be applied within ML. Algorithms are set of rules for solving a problem in a finite number of steps. Alternatively, as Mahmud et al. (2022) defined:

"An algorithm is a set of encoded procedures for transforming input data into a desired output based on specified calculations. It is an automated process that provides decisions independently without the mediation of a human. Algorithms employ data, statistics and computing resources to make decisions." – (Mahmud et al., 2022, p.2)

Table 3

Algorithms in Machine Learning Models adapted from (Akter et al., 2022)

Supervised Learning	Unsupervised Learning
Classification	**Clustering**
- Neural Networks - Decision Tree	- Hierarchical Clustering - K-Means
Regression	**Dimension Reduction**
- Linear Regression - Neural Network	- Principal Component Analysis - Singular Value Decomposition

There is a difference between algorithms and AI. An algorithm operates as instructed, whereas AI is coded to learn how to perform an operation and modify its programming if given new data (Buiten, 2019). Contrary to an algorithm, an AI system allows a machine or program to improve its performance by learning (itself) from its given dataset. The more training data an algorithm is provided, the more precise and accurate the outcomes will be (Zhang & Lu, 2021). This is illustrated in the following steps, which describe the process of training an AI system. Based on Kleinberg et al. (2018) and van Giffen et al. (2022) the process of training AI includes the following steps:

1. The AI developer selects a dataset based on existing data, known as 'historical data'.
2. The AI developer decides the desired output is applying this dataset; What is the final decision that has to be made?
3. The AI developer decides which instructions to give the algorithm. These instructions are based on the AI system's end goal; Which predictions can an algorithm make?
4. The AI developer decides on the best outcome of the AI system, identifying all variables that need to be evaluated; What does the ideal outcome look like?
5. The algorithm will be tested with data other than the dataset employed to build it. Testing the algorithm verifies the benchmark of the dataset it trained initially with regarding the representativeness of all subgroups. It reveals if some subgroups are more represented than others.

This thesis uses the CRISP-DM Model (Cross Industry Standard Process for Data Mining) to evaluate how an AI system develops and operates. Based on this model, biases within AI will be identified, and mitigation methods for these biases will be developed. The CRISDP-DM model clarifies an AI developer's steps and choices in designing an ML model. Since 1999 it has been the most common process model for data mining, data analytics and data science projects (Bechmann & Bowker, 2019; van Giffen et al., 2022). The model is also employed in studies to display an AI system when the 'real' AI system is unknown. In particular, ad delivery platforms identify their AI systems as their business model and will not share it with the outside world (Bechmann & Bowker, 2019; Rijksoverheid, 2021). This thesis copies the CRISP-DM Model design developed by van Giffen et al. (2022), see Figure 2, which is different from the original method of the CRISP-DM Model because it also describes the potential biases that might occur during the process.
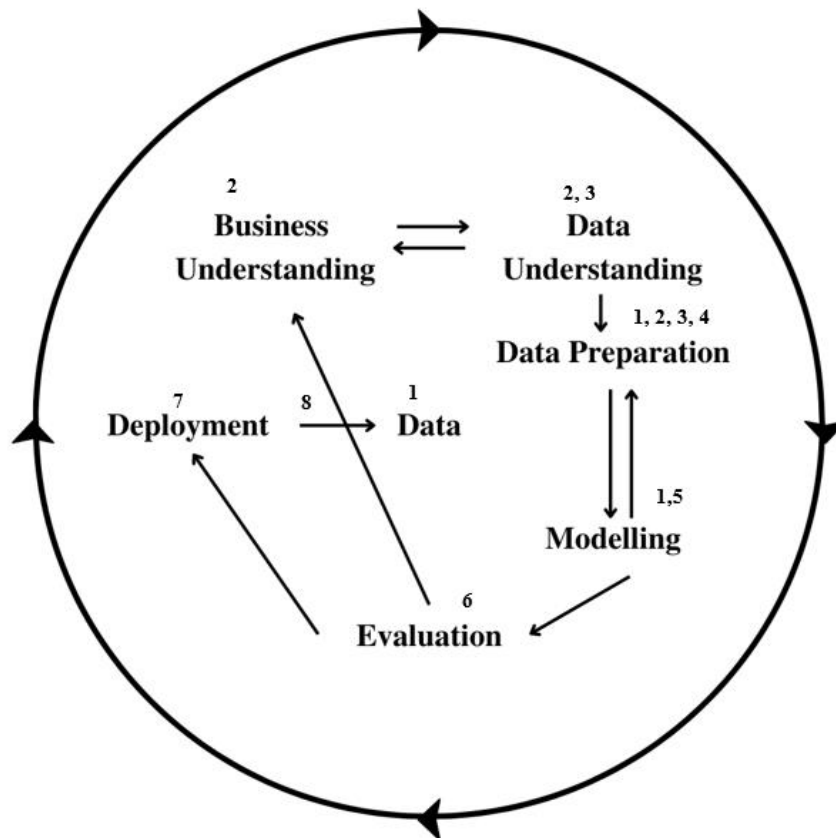
Figure 1

CRISP-DM Model adapted from (van Giffen et al., 2022)

The CRISP-DM Model is structured as follows: The 'Business Understanding' phase focuses on understanding the objectives and requirements of the AI project. This knowledge is then converted into an AI problem definition. Next, a plan is designed to reach these objectives. The 'Data Understanding' phase begins with the gathering of data. It continues with activities that enable the developer to become familiar with the data, such as: describing –, exploring –, and assessing data quality. The activities required to create the final dataset from the initial raw data are included in the 'Data Preparation' phase. These activities include observing, selecting attributes, transforming and/or cleaning the data. The final dataset will be administered to train an ML model. In the 'Modelling' phase, several ML techniques are selected, which will then be applied to the dataset. The 'Evaluation' phase reviews the outcomes of the ML techniques and decides which approach meets the success criteria. To ensure that the ML model successfully meets its objective, it is crucial to evaluate – and examine the processes employed to generate it. The final 'Deployment' phase involves implementing the (self-learning) algorithm for organisational decisions to reach the defined objectives from the 'Business Understanding' phase.

1.3. How do Biases occur?

The word 'bias' is a synonym for preferring one person or thing to another person or thing (Akter et al., 2022). In the Cambridge dictionary, bias is found to be synonymous with the words 'favouritism' and 'prejudgement'. When AI systems are biased, the systems make decisions or predictions favouring one thing above another and are identified as 'unfair' (Cambridge Dictionary, 2022). As previously discussed, several studies agreed that AI could lead to algorithmic biases, which can then lead to harmful unintended outcomes (Akter et al., 2022; Ali et al., 2019; Bandy, 2021; van Giffen et al., 2022).

According to Akter et al. (2022) and van Giffen (2022), biases can be caused by unrepresentative datasets, inadequate models, weak algorithmic design or human stereotypes that result in potentially harmful unintended outcomes. These causes differ from the assumptions made by the writer Cathy O'Neil who argues in her book 'Weapons of Mass Destruction' that unintended outcomes occur due to deliberate or unconscious biases by humans who program the algorithm. Alternatively, the algorithm has learned how to be biased based on the behavioural data it taught (O'Neil, 2017).

Ali et al. (2019) are against the fact that biases are intentionally implemented and argue: *"...discrimination may be introduced by the ad platform itself, rather than by a malicious advertiser."* – (Ali et al., 2019, p.9)

Ali et al. (2019) researched how ad delivery systems provided by ad platforms could lead to biased outcomes. This study redesigned Facebook's algorithms and proved that Facebook's previously unknown AI system could lead to unintended outcomes, specialising in discriminatory ad delivery.

In line with the CRISP-DM model from van Giffen et al. (2022), this thesis has created examples for each bias category in Table 4. It will go into further detail about specific biases in Chapter 2. Additionally, this thesis will investigate possible mitigation methods to prevent these biases from occurring.

Table 4

Type of Biases (Own elaboration)

Type of Bias	Bias definition	Real-life example
1. **Social Bias**	Social bias occurs when the available data reflects the existing biases in the population before developing the algorithmic AI model (Olteanu, Castillo et al., 2019).	Algorithms can be trained with country-specific data. This country-specific data reflects the differences in gender roles in a country's culture. Over time, the algorithm was trained with this country-specific data and has learned to reflect this bias in its outcomes. When the same algorithm is employed in a different country with a different perspective on gender roles, this pattern of social bias or gender discrimination is revealed (Lambrecht & Tucker, 2019).
2. **Measurement Bias**	Measurement bias occurs when the labels that are chosen to design the algorithmic model are imperfect for the actual variables of interest (van Giffen et al., 2022)	Tay et al. (2022) identified that applying models that contained measurement biases in automated assessments of job interviews may disproportionately exclude minorities from a job.
3. **Representation Bias**	Representation bias occurs when the input data is not representative of the relevant population and will therefore lead to systematic unintended outcomes (Olteanu, Castillo, et al., 2019; van Giffen et al., 2022).	Twitter's image-cropping algorithm automatically focused on white faces instead of on black faces. Twitter offers a feature that uses an algorithm to focus on the most critical parts of the image. While doing so, it does not recognise black-coloured faces. The reason for this "racist" image cropping algorithm is that the algorithm was trained by a dataset that contained more data about light-skinned people than dark-skinned people. In conclusion, the dataset did not represent the relevant population (all skin-coloured Twitter users) (Blansett, 2019; Hern, 2020; Lambrecht & Tucker, 2019).

4. Label Bias	Label bias occurs when the training data is assigned to the wrong class labels (Olteanu, Castillo, et al., 2019; van Giffen et al., 2022)	A class label has been changed, but the newly generated one does not match the newly created class label and the historical data that serves as input to the machine learning model (van Giffen et al., 2022).
5. Algorithmic Bias	Algorithmic bias occurs when the model systematically deviates from the outcome due to inappropriate technical considerations in the modelling phase (Ann et al., 2022; van Giffen et al., 2022).	When a bank develops an algorithm for mortgage customers, it evaluates multiple demographic and financial variables. The nature and weight of each of these variables could lead to a deviated outcome and bias against certain customer groups (Akter et al., 2022a).
6. Evaluation Bias	Evaluation bias occurs when a non-representative population is administered to evaluate the model (Olteanu, Castillo, et al., 2019; van Giffen et al., 2022).	The algorithmic model employed for an online advertisement is trained and optimised based on the user's data of the company. However, when the company would like to test and evaluate the advertisement, the company uses a different dataset than the algorithmic model is trained for (van Giffen et al., 2022).
7. Deployment Bias	Deployment bias occurs when the model is interpreted in a different context than it was built for (Olteanu, Castillo, et al., 2019; van Giffen et al., 2022).	An algorithm model employed in revenue management to assist with pricing and decisions may only be effective for some types of products. The same algorithm may produce biased results when applied to products from a different context (De-Arteaga et al., 2022).

| 8. **Feedback Bias** | Feedback bias occurs when the model's output is employed as new inputs for the model. Therefore, biased outcomes are reinforced within the feedback loop (Olteanu, Castillo, et al., 2019; van Giffen et al., 2022). | According to Ann et al. (2022), police forces apply racial profiling to predict crimes. Part of racial profiling is the higher number of surveillance routines done by police forces in racially profiled neighbourhoods, leading to more crimes being registered. An algorithm derived from datasets established on racial profiling would reinforce the pattern that in racially profiled neighbourhoods, there are more crimes. This would likely start a cycle and lead to feedback biases. |

1.4. Which Laws and Regulations are proposed and adopted by the European Commission?

AI is increasingly prevalent due to rising digitalisation (Bringas Colmenarejo et al., 2022; Ngai & Wu, 2022). However, existing laws and regulations lack governance mechanisms to mitigate the risk of harmful unintended outcomes in applying AI (Chhillar & Aguilera, 2022; European Digital Rights, 2021). Therefore, the expanding employment of AI has motivated the European Commission (EC) to develop the EU Artificial Intelligence Regulation Proposal (AI Act), the first law on AI worldwide (Bringas Colmenarejo et al., 2022). The EC acknowledges the importance of the development of AI and the competitive advantages it brings to companies and the European economy. However, it argues that the power of AI that can lead to unintended outcomes, which can be harmful to individuals or society. Thus they should be regulated (European Commission, 2021).

Before discussing the AI Act, it is essential to understand that the EC has provided other laws and regulations. The AI Act is one of the several proposed laws that must be considered. Veale & Borgesius (2021) provided an overview of all the laws and regulations adopted and proposed by the EC.

Table 5

Laws and regulations adapted from (Veale & Borgesius, 2021)

Law/Regulation	Status
General Data Protection Regulation	Adopted on the 14th of April 2016
Data Governance Act	Proposed on the 21st of February 2021
Digital Markets Act	Adopted on the 25th of March 2022
Digital Services Act	Adopted on the 23rd of April 2022
Product Liability Revision	Adopted on the 28th of September 2022
AI Act	Proposed on the 21st of April 2021
Machinery Regulation	Proposed on the 21st of April 2021

To remain the focus of this research subject: The viewpoint of companies using AI in their business practises on the AI Act, this thesis will not go over all these laws and regulations in detail. The AI Act will remain the primary focus, and thus the law that is covered in its entirety. On the other hand, this thesis will examine how laws and regulations prevent the biases identified in Table 4.

1.4.1. The AI Act

The EC was in 2021 the first governmental institution to issue an EU Regulation Proposal, also known as 'the AI Act' or 'the AI Act Package'. This was done in response to the broadening application of AI in business practices, thus expanding demand for AI regulations. The law focuses on providers, users, agents, importers, and distributors responsible for developing and deploying AI. The EC defines the following objectives of the law within its white paper on AI:

1. *"Ensure that AI systems placed on the Union market are safe and respect existing laws on fundamental rights and Union values.*
2. *Ensure legal certainty to facilitate investment and innovation in AI.*
3. *Enhance governance and effective enforcement of existing law and safety requirements applicable to AI systems.*
4. *Facilitate the development of a single market for lawful, safe, and trustworthy AI applications preventing market segmentation."* – (European Commission, 2021, p.3)

The AI Act states that AI systems will not be prohibited from entering the EU market if they meet the requirements outlined in the legislation. Table 6 presents how these requirements differ depending on the AI model's risk level. The AI Act classifies AI systems as products. As a result, the Product Liability Directive is incorporated into the AI Act. People can seek compensation under the Product Liability Directive if an AI system is defective or fails to perform as expected (for example, in online advertising) (European Commission, 2021).

Table 6

Updated Risk categories AI Act (Own elaboration)

Unacceptable risk AI systems	High risk AI systems	Limited – and minimal-risk AI systems
Prohibited	**Restricted**	**Allowed**
Subliminal -, manipulative – or exploitive techniques causing harm Example: Targeting vulnerable groups (children or disabled people) to disturb them better than is likely to cause them or somebody else psychological – or physical harm (Neuwirth, 2022b).	**Systems that evaluate customers' creditworthiness** Example: A bank uses an AI system to determine whether a customer can borrow money (Benjamin et al., 2021).	**AI chatbots** Example: applying an AI system to communicate at a human level with customers through the help of a text-based dialogue system (Huang & Rust, 2021; Ngai & Wu, 2022).
Real-time, remote biometric identification systems are applied in public spaces or law enforcement Example: Police tracking people on the street or in governmental buildings based on specific characteristics, such as skin colour, hair, and ethnicity (Bessen et al., 2022; Neuwirth, 2022b)	**Biometric identification in non-public spaces** Example: An company keeps a biometric categorisation system that keeps track of its customers' sexual or political orientation (Benjamin et al., 2021; Neuwirth, 2022b).	**Customer – and market-segmentation systems** Example: Applying AI systems to identify patterns and create customer segments by dividing a market into segments based on demographic, geographic, and socio-graphic factors (Huang & Rust, 2021; Ngai & Wu, 2022).
All forms of social scoring Example: A public authority keeps a social score of its citizens for a general purpose (Neuwirth, 2022b).	**Recruiting – or employee management systems** Example: An company uses an AI system based on the resume of job applicants to determine who is applicable for a job and who is not (Benjamin et al., 2021).	**Spam filters** Example: Teaching an email service which emails are identified as spam and which are not (Akter et al., 2022).

Unacceptable risks AI systems will no longer be permitted within the EU. High-risk AI systems will be subject to new requirements – including human oversight, transparency, implementation of a risk-management system, data governance and management, technical documentation, record keeping and logging, accuracy, robustness, conformity assessment, registration with EU-member-state government, post-market monitoring procedures to minimise the risk of algorithmic biases. Limited – and minimal risk AI-systems will most

likely not be subject to this new AI Act proposal but will still need to meet the requirements of the GDPR.

According to the AI Act, every company should enhance a three-action comprehensive AI-risk management program integrated within its business operations: 1) create an inventory of all its AI Systems and risk-mitigation measures, 2) conduct conformity assessments, and 3) establish an AI governance. The three actions are based on a feedback loop in which each activity will inform and improve the following activity.
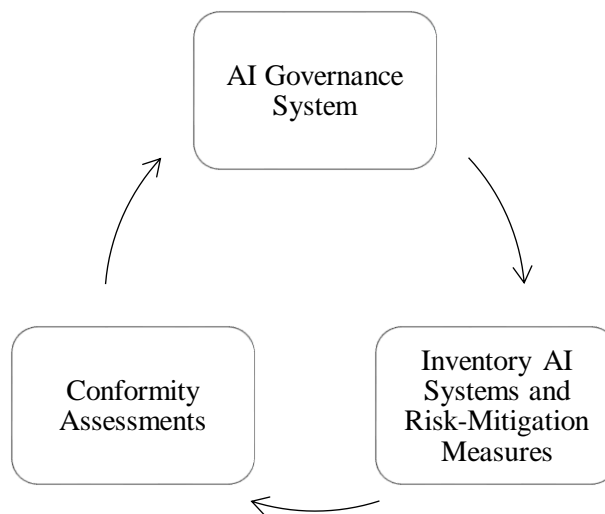

Figure 2

Improved organisational AI-risk Management Program adapted from (Benjamin et al., 2021)

AI Governance Systems

Companies should form a committee of professionals from various functions to ensure the responsible development and deployment of AI (ethics, legal, cyber security, product development, etc.). These professionals are in charge of establishing the standards that the AI system must meet to be used in companies and advising business and development teams.

Inventory of AI Systems and risk-mitigation measures

Companies should create and maintain inventories that describe which AI systems are used and which risk classifications apply to each AI system. Mapping AI systems against existing and potential future regulations enable the discovery of compliance gaps. Furthermore, a legal person can rely on this inventory to create a standardised risk classification taxonomy compliant with current and future AI laws and regulations. Finally, this measurement identifies high-risk AI systems and identifies mitigation methods for them.

Conformity Assessments

A conformity assessment reviews each AI system to determine whether it complies with current and upcoming AI regulations. Documentation of the various choices made when developing AI systems, as well as their associated limitations and level of accuracy; an overview of the risks that the AI system carries, including the potential sources of unintended outcomes; and any risks associated with the AI system, as well as mitigation methods to mitigate these risks. Conformity assessments enable companies to report on their processes, risk evaluation, and mitigation, all while ensuring AI developers use best practises. The conformity assessment will generate a report on the company's process as it moves through the various risk evaluation and mitigation stages, including data evaluation to address biases (Benjamin et al., 2021; Bringas Colmenarejo et al., 2022; Neuwirth, 2022a).

To better understand the advantages of the implications of the AI Act, this thesis quotes Werner Stengg, a member of the EU cabinet. Werner Stengg argues why a proposal on regulating AI is important and shared his expertise in an online conference on the EU AI Act organised by Stanford Human-Centered Artificial Intelligence.

"In the past, governments have established guard roll laws and norms to balance innovation and protection, but these laws and norms have not been updated to apply when machines and data perform the activity." – EU AI Act - Stanford Human-Centered AI Online Conference - Werner Stengg discourse. 2021, 2:23

He continued his arguments at the conference by providing the following three arguments for why the EC should regulate AI:

1. Democratic Legitimacy to companies applying AI
   Technology brings advantages to the economy and society. However, it has taken a turn that it has yet to be democratically decided what type of internet we want and all the societal and economic effects that come with it.

2. Relationship innovation and regulation
   It is a myth that regulations will stifle the development of AI. There is no contradiction between innovation and regulation. If AI is appropriately applied, it can do wonders. However, if it is not correctly applied, it can put fundamental rights at risk.

3. Regulation not upon private companies

   Governors cannot rely on private companies to police themselves. The main goal of companies is to create profit. It is the government's job that this can be done safely by creating regulations.

However, according to experts, the new regulation contains several loopholes and exceptions. Borsci et al. (2022), Bringas Colmenarejo et al. (2022) and Veale and Borgesius (2021) reviewed the AI Act and argued why some implications are ineffective and need improvement, see Table 7. These inabilities limit the AI Act's ability to ensure that AI can be run safely in the long term. In addition, the AI Act is inflexible. If a high-risk AI system application is applied in an unanticipated sector, the AI Act provides no mechanism to identify this system as 'high risk'. Because the AI Act cannot identify the sector risk, new high-risk AI system applications can be introduced into the market, leading to conflicts.

Table 7

Improved organisational AI-risk Management Program (Own elaboration)

Name Organisation	Suggestion of improvement
Future Life Institute	AI providers should evaluate the impact of their AI applications at large and not only focus on the impact of an individual (Future of Life Institute, 2021).
University of Cambridge Institutions	Changes to the unacceptable – and high-risk AI systems should be allowed to increase the flexibility of the AI Act (University of Cambridge, 2021).
Access now Europe	The AI Act is unable to reach the objective of protecting fundamental rights and should therefore be stricter (Access Now Europe, 2021).
Michael Vaele & Zuiderveen Borgesius	A weakness of the AI Act is that the regulation would now almost rely entirely on self-assessment. See Figure 2. This means that third-party verification with the law will no longer be required. Therefore, Veale & Borgesius (2021) advocate for establishing third-party verification.
The Future Society	The AI Act should remain responsive to technological trends. This can be reached by improving the information stream between national and European institutions. In addition, incident reports (where AI has led to – harmful – unintended outcomes) should systematically be gathered and evaluated (The Future Society, 2019).
Nathalie Smuha & Colleagues	The AI Act does not frequently recognise the wrongs and harms associated with different AI systems. In addition, the AI Act is ineffective because it neglects to ensure transparency, accountability and public participation rights (Smuha et al., 2021).
The European Digital SME Alliance	The AI Act can be interpreted to be impractical for SMEs. To prevent this from happening, SMEs should get an active role in developing the standards of the AI Act, such as developing standards for conformity assessments (European DIGITAL SME Alliance, 2022).

1.5. Which Laws and Regulations focus on which Biases?

Table 8 concludes which laws and regulations prevent the biases identified in Table 4 and if any potential blind spots are not addressed.

Table 8

Regulations and Biases (Own elaboration)

Regulation	Date	Type of Bias identified from the CRISP-DM Model
General Data Protection Regulation	Adopted on the 25th of May, 2018 (European Parliament and Council of the European Union, 2016; Local Government Association, 2018)	Social -, Representation Bias
Data Governance Act	Proposed on the 21st of February 2021 (European Commission, 2022)	Social -, Representation Bias
The AI Act	Proposed on the 21st of April, 2021 (European Commission, 2021c)	Social -, Measurement -, Representation -, Label -, Algorithmic -, Evaluation -, Deployment – and Feedback Bias (depending on which level or risk the AI system falls)
Machinery Directive	Proposed on the 21st of April, 2021 (European Parliament, 2022)	Social -, Measurement - , Representation -, Label -, Algorithmic -, Evaluation -, Deployment – and Feedback Bias
Digital Markets Act	Adopted on the 25th of March 2022 (European Commission, 2022)	Focusses on how companies employ AI. Does not focus on the AI system itself.
Digital Services Act	Adopted on the 23rd of April 2022 (European Commission, 2022)	Focusses on how companies employ AI. Does not focus on the AI system itself.
Product Liability Directive	Adopted on the 28th of September, 2022 (European Commission, 2021)	Social -, Measurement - , Representation -, Label -, Algorithmic -, Evaluation -, Deployment – and Feedback Bias

Laws and regulations highlight all the possible biases. According to the findings of Table 8, the AI Act is one of the most robust regulations. Along with the Machinery - and Product Liability Directive, it focuses on all biases. In contrary to the Digital Markets – and Services

Acts, which do not assess any biases caused by AI. The described biases only apply to AI, and the laws and regulations also apply to Big Data and ICT. Social and measurement biases are the most well-protected biases. Both biases represent those that could occur in an AI system's data. The deployment bias is the least protected bias.

2. STUDY 1: CASE STUDY

A case study is included within this thesis to answer the research questions. The objective of this case study is to examine how the ad platform Facebook uses AI in its advertising and how this can lead to biases. These findings provide an answer to the research question: How are businesses incorporating Artificial Intelligence into their advertising campaigns?

2.1. Research Design

A case study is a method to describe and evaluate a research problem (McCombes, 2022)**.** The value of a case study is recognised in several fields of business, law and policy and allows the researcher to obtain an in-depth understanding of an issue in its real-life context (Crowe et al., 2011).

The case study within this thesis has been established by applying secondary sources (Scribbr, 2018). In addition, documents, legislation, articles and Facebook's policies have been analysed. This thesis focuses on how Facebook delivers ads as a case study subject, as it is one of the most influential and feature-rich advertising platforms currently employed   (Ali et al., 2019; Dalenberg, 2018).

2.2. Results

Facebook lets other companies employ their platform to target their users with advertisements. Additionally, Facebook offers banner advertising space on various websites and mobile apps (Dalenberg, 2018). As a result, the ad platform delivers advertisements to billions of users daily.

This thesis defines the functionality of Facebook as an advertising platform in the following two phases. One is the ad creation phase, where advertisers can design their ads by entering a combination of text and images. The second is the ad delivery phase, where Facebook delivers the ads to the target audience based on a variety of the performance of an ad, the predicted relevance and the advertiser's budget.

According to Facebook Business (2015) and Ali et al. (2019), during the ad creation process, advertisers need to make three choices before moving onto the ad delivery phase:

1. Who is the desired audience?
   The desired audience is determined based on the user's previous activities on Facebook, such as liking Facebook pages or interacting with ads and the user's information directly shared with Facebook, such as gender and age. Additionally,

advertisers can share details about their existing customers or previous Facebook users who have already interacted with their website or Facebook Page, also called: the look-a-like audience. (Facebook, 2020; Facebook Business, 2015)

2. What is the advertiser's objective?
   The advertiser determines the objective of the advertisement. Facebook allows the advertiser to choose between three objectives 1) Awareness, creating brand awareness and increasing reach; 2) Consideration, increasing traffic and customer engagement. Lastly, 3) Conversion, creating conversions.

3. How much is the advertiser willing to pay per ad?
   Advertisers must state their price range to have their adverts displayed. By placing a per-impression or per-click bid, advertisers can set a bid amount cap which allows Facebook to submit bids on the advertiser's behalf.

After creating an advertisement, Facebook reviews the ad to see if it matches its policies, after which the ad will be moved to the ad delivery phase. Within the ad's delivery phase, the ad slot and an ad's visibility per user will be decided.

The decision of which ad will be presented to which person will be decided through a combination of the ad auctions and the 'Total Value' of an ad.

In the auction stage, advertisers set a bid indicating how much the marketing agency is prepared to pay for a Facebook user to interact with their advertisement and achieve their goal. Facebook's ad auction algorithm then chooses the most effective advertisements for each ad impression based on the advertisers' maximum bids and ad performance. In this procedure, all advertisers compete against one another, and the ads that Facebook determines to have the highest chance of success will take the top spot.

Through the help of its AI model, Facebook will calculate the 'Total Value' of the ad. The total value decides who gets to interact with the ad is computed through the following equation:

$$Advertiser's\ bid\ x\ Estimated\ Action\ Rate + Ad\ Quality = Total\ Value$$

Equation 1: Total Value

Through ML, Facebook estimates the 'Action Rate'. The Action Rate indicates how likely a Facebook user will complete the advertiser's objective. The probability of reaching this objective is decided based on the following factors:

1. How active is someone on Facebook? E.g., engaging with the content on their newsfeed.
2. What are the user's off-Facebook activities? e.g., interacting with the company's website.
3. How are similar profiles, and look-a-like audiences, performing?

Finally, Facebook uses its AI system to judge 'Ad Quality and Relevance Diagnostics'. The Quality and Relevance Diagnostics are calculated based on the positive – and negative Feedback expected an ad receives from its target audience. Relevance Diagnostics is graded with a score ranging between 1 through 10 (Facebook Business, 2015).

The total value depends not only on the advertiser's bid but also on how likely users are to interact with the ad and the ad's quality. The equation allows SMEs to benefit from Facebook's ad policies and helps them promote their company (Facebook, 2020; Facebook Business, 2015). Alternatively, as Facebook itself puts it:

"Put simply, the higher an ad's relevance score, the less it will cost to be delivered. This is because our ad delivery system is designed to show the right content to the right people, and the system sees a high relevance score as a positive signal" – (Lambrecht & Tucker, 2019a, p.2969)

The following chapters are a review of the biases that arise within Facebook's advertising system with a focus limited to Social – and Algorithmic Bias. This limitation is a consequence of the lack of insight into Facebook's delivery algorithm, - statistics and user – and targeting data.

2.2.1.   Social Bias

Facebook identifies which ads particular users are most likely to engage with before the advertisers – who pay Facebook to display their ads – have identified these users themselves (Blass, 2019).

Facebook: *"We try to show people the ads that are most pertinent to them."* – (Ali et al., 2019, p.2)

To achieve this, ad platforms like Facebook have developed in-depth user interest profiles and track the performance of advertisements to understand how a different target audience interacts differently with an ad. This data is gathered as 'historical data' and is then used to target future advertisements to target audience and look-a-like audiences who are most likely to be interested in them. This historical data can consist of likes, and third-party data, based on behaviour from lookalike audiences, which help Facebook predict ads a user might be interested in (Akter et al., 2022; Andreou et al., 2019; Bashir et al., 2019).

However, Facebook might unintentionally create social biases when applying this historical data (Ali et al., 2019; Kingsley et al., 2020). This means that advertisements can be delivered to a skewed subgroup of the advertiser's selected target audience. A potential cause for social biases is if the biases of the people who created these datasets are embedded in them, AI systems trained on those datasets are likely to repeat existing biases or even introduce new biases based on patterns that happen to be present in those data. However, such biases may go largely undetected because machine learning algorithms are challenging to check or understand (Blass, 2019; O'Neil, 2017).

Ali et al. (2019) found evidence of social biases due to AI-driven online campaigning by finding evidence of stereotypical online advertising outcomes. These results were reached by running ads on Facebook to measure how the ad delivery differs along racial – and gender lines. Ali et al. (2019) designed stereotypical ads focused on gender. One ad targeted to male users was concentrated on bodybuilding. The ad targeted to female users was focused on cosmetics. When they placed the same advertising budget on the same target audience, the advertisement's delivery still led to stereotypical outcomes due to the historical data Facebook's ad delivery system has been trained with. Ali et al. (2019) detected that ads related to bodybuilding were delivered to over 75% of men, and ads about cosmetics were delivered to 90% of women.

These biased outcomes were repeated when a similar experiment employed ads focused on the impact of cultural content on the ad delivery along racial lines. The ad employed to target black users focused on hip-hop music, and the ad employed to target white users were focused on country music. Without specifying the advertising budget and target, the target audience of hip-hop advertisements consisted of 85% of black users. At the same time, the target audience of country music was 80% of white users.

Because Facebook's delivery algorithm is trained with historical data, it implements historical biases. Historical Bias is a synonym for Social Bias (van Giffen et al., 2022). The biases from the dataset are an example of indirect discrimination, which cause extreme differences in the outcomes of the ad delivery. Indirect discrimination occurs when an advertisement's neutral settings exclude a significant portion of its target audience, as is presented in the examples (Dalenberg, 2018; van Giffen et al., 2022).

2.2.2.   Algorithmic Bias

Lambrecht & Tucker (2019) found evidence that applying algorithms in online advertising on Facebook might lead to unintended outcomes. This finding was reached in their research on gender biases in STEM (Science, Technology, Engineering, Mathematics) careers.

There is a chronic shortage of female graduates going into STEM fields worldwide (VEX Robotics, 2017). In the United States, only one in seven females is an engineer; in the United Kingdom, only 6% of the STEM workforce is female (CNN, 2014). To learn how to attract more female graduates into the STEM workforce, research was performed on how men and women were recruited through online advertisements for STEM careers. Through a field research, Facebook ads were distributed to men and women between 18 and 65 years old from 191 countries. This research showed that 20% more men than women saw the Facebook ad and that women between 25 and 34 years old were 40% less likely to interact with the ad than men of the same age (Lambrecht & Tucker, 2019b).

The reason for these biased results is that Facebook's algorithm unintendedly excluded women from interacting with the STEM Career ad because this favoured the advertising auction (Bessen et al., 2022). Facebook has been criticised for implementing cost optimisation strategies that focus on the revenue per click and return investment, disregarding their ethical and legal obligations (Ali et al., 2019). An algorithm trained to be cost-effective will automatically exclude the more expensive option, in this case, female users. Algorithms can be a 'black box'. Therefore, advertisers do not always understand the features that the AI model uses to create an outcome. Companies applying AI may not always be aware of the causes of potential discriminatory decisions (Ali et al., 2019). This case study continues to explain why and how, in Facebook's example, an algorithm is optimised to make cost-effective decisions that can sometimes be discriminatory.

Online advertisers compete to display their ads to the same set of 'eyeballs'. This competition can influence how ads are distributed, even if the ads promote different products

and/or services. Literature illustrates that based on the general distribution of ads, it can be concluded that women are more likely to employ social media than men, are more active on social platforms and are consequently more likely to be exposed to ads. Additionally, literature proves that women are generally more likely to convert and purchase after clicking an advertisement than men (Akter et al., 2022).

Lambrecht & Tucker (2019a) support the statement that women are principally more expensive to target. Lambrecht & Tucker (2019a) illustrate this statement by sharing statistics of the average bid suggested by Facebook to reach female users specifically. Their results illustrate that the average suggested bid to target a Facebook user, regardless of gender, is $0.45. At the same time, the average bid to target female users is $0.50. In addition, Lambrecht & Tucker (2019a) replicate the STEM Advertising campaigns on Instagram. Facebook and Instagram are owned by Meta and share a similar advertising platform (Meta, 2023). These results illustrate that the cost per click for a female user is $1.74, whereas the cost per click for a male user is $0.95. This difference in cost per click led to 1,560 female users being able to interact with the STEM job advertisement and 9,595 male users.

In summary, female users are more expensive to advertisers than male users. Advertisers who bid with a low budget when placing their ads may be more likely to lose auctions for such "expensive" users than those with a high budget. However, suppose that the expensive user demographics strongly correlate with users from a specific gender or ethnicity and, therefore, more to target. In that instance, it could lead to a discriminatory ad delivery due to the advertiser alone. Even though an advertiser with a low budget does not intend to exclude such users, Facebook's ad delivery system may do that (Ali et al., 2019; Blass, 2019).

The gender imbalance of the STEM ads reflects that women are more likely to convert after interacting with an ad and, consequently, are more expensive to present ads to. The algorithm optimises the ad delivery to be cost-effective and delivers ads intended to be gender-neutral in an accidentally discriminatory way. In conclusion, it is neither Facebook nor the companies applying Facebook to promote their advertisements objective to create biased outcomes. It is the algorithm performing the job it is trained to do: Saving companies money by delivering the ad against the lowest price (Žliobaitė, 2017). Therefore, this example is considered to be an Algorithmic Bias. The examples supporting the type of biases prove that unintended outcomes within online advertising are not intentionally caused by an

advertiser preferring one target audience over the other. The models conclude that biases within online advertising occur when ad platforms try to deliver the ads to their users.

3. STUDY 2: INTERVIEWS

The second section of the results focuses on semi-structured interviews with marketing –, AI – and legal experts. The interviews aim to answer the research questions and gain insight into the application of AI within online advertising.

3.1. Research Design

This thesis uses a qualitative approach to provide more insights into the complexity of the subject: 'Unintended Outcomes due to the Application of Artificial Intelligence. The interviews have been accounted for as a primary source. Because the outcome of this thesis is a legal framework, which takes a snapshot of the current affairs regarding AI regulations, this thesis is descriptive (Hoover, 2021; Scribbr, 2018).

This thesis started with gathering data from 25 interviewees, in line with the advisable 5 through 50 interviewees, to design trustworthy qualitative research (Dworkin, 2012). However, after contacting 70 experts and companies, this number was reduced to 18 interviewees during the writing of this thesis due to a lack of correspondence and collaboration. The lack of participation of interviewees has led to a non-participation ratio of 74.3%. The final list of interviewees contributing to this thesis has been included in the appendices of this thesis.

Interviewees were included in this thesis because of their expertise in one or more of the following areas: They were applying AI within their online marketing activities (Marketing), developers of AI systems (Technology), and interviewees were active in the legal or legal tech field legal (Legal). Interviewees have been identified based on desk research, snowballing, and participation in the Dutch "National AI Course" and "National AI Course on Ethics". Potential interviewee candidates were contacted and invited for an interview through e-mail, telephone, LinkedIn and/or through the company with which the interviewee was active. A demographic profile of the anonymised interviewees has been created; see Table 9.

Table 9

Demographic Analysis Interviewees (Own elaboration)

Demographic Characteristics	Sub – Level	Count (n=18)	%
Gender	Male	9	50%
	Female	9	50%
Age	25 – 34	7	38.9%
	35 – 44	4	22.2%
	45 – 54	5	27.8%
	55 – 64	2	11.1%
Company size	Small company with 10 – 99 employees	6	33.3%
	Medium company 100 – 499 employees	2	11.1%
	A large company with 500 employees	10	55.6%
Position	CEO	1	5.6%
	Digital Analytics Expert	3	16.7%
	Lawyer	4	22.2%
	Marketeer	3	16.7%
	Professor	2	11.1%
	Researcher	4	22.2%
Industry	Educational Services	8	44.4%
	Information Technology	4	22.2%
	Marketing	3	16.7%
	Public Services	3	16.7%
Expertise	Use of AI in Marketing	6	33.3%
	Technological part of AI	5	27.8%
	Technological part of AI / Legal Regulations of AI	3	16.7%
	Legal Regulations of AI	4	22.2%

The interview questions shared with the interviewees have been listed in the appendices of this thesis. Interviews have been held in a semi-structured way. This means that the interview questions were the main starting point of the interview but that when needed, additional and follow-up questions were asked as well (Tegan, 2022).

To evaluate the qualitative data, the Gioia Method has been applied. This method develops first-order codes (information-centric terms) and second-order codes (theory-centric themes) into overarching theoretical dimensions. Finally, these developed the gathered dimensions, and these codes created a data structure that helped answer the research questions (Dennis A. Gioia, 1990).

In the second part of the results Chapter, the output retrieved from interviewees will be outlined by focusing on the three sections: Marketing, Technology and Legal. In the marketing section, the application of AI within advertising activities is investigated by creating an overview of the activities necessary to develop an online ad. The technological section outlines the causes of intended outcomes and potential mitigation methods. This Chapter is finalised with the legal section, which evaluates which legal frameworks exist to mitigate these causes of unintended outcomes. Finally, a list of check questions is presented, which can help companies mitigate the causes of unintended outcomes in online advertising during the development of an AI system.

3.2. Results

The second section of this Chapter is divided based on the three main questions focusing on the strengths, weaknesses, opportunities, and threats of marketing agencies applying AI (Marketing), the causes and prevention of unintended outcomes within AI (Technology) and the laws and regulations regarding AI legal.

3.2.1. Marketing Agencies' approach to applying Artificial Intelligence in Online Advertising

The interviewees suggest in which different ways marketing agencies in the field of online advertising apply AI. The findings from these interviews have led to the following results presented in Figure 3.
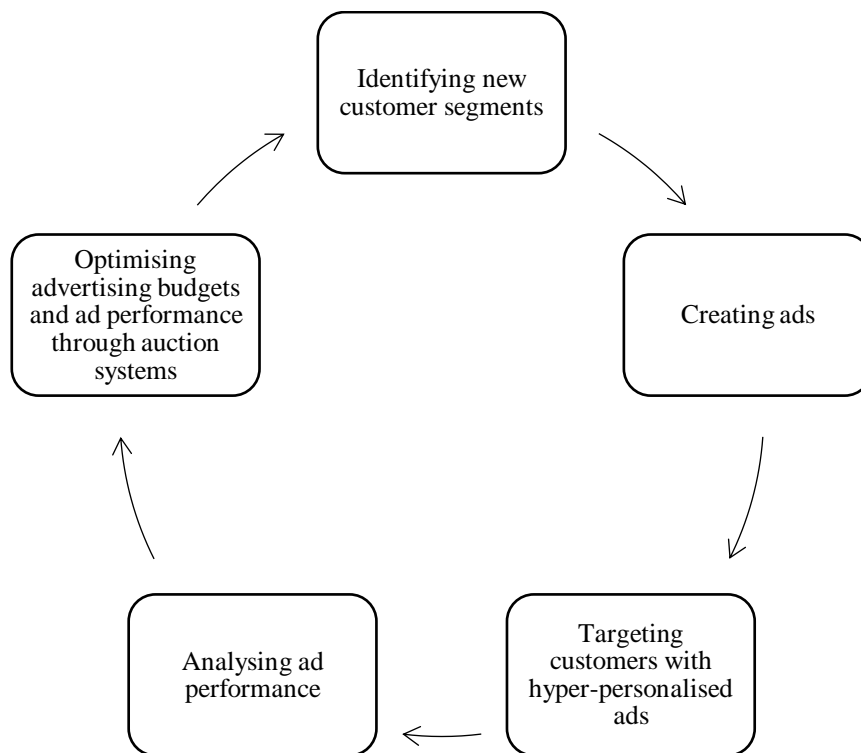

Figure 3

Application of Artificial Intelligence in Online Advertising (Own elaboration)

Identifying new customer segments

It is critical to identify the target audience the marketing agency wishes to target before creating ads. It is possible to identify new customer segments using AI systems. A marketing agency that participated in the interviews shared an example of a customer who initially identified the target demographic of its newly launched product as 65 and older. An experiment in which an AI system could target not only older adults but also potential

customers between 18 and 35 resulted in the conclusion that the younger target group between 18 and 35 was far more profitable. This example demonstrates that the AI system evaluated data and concluded that the target group between 18 and 35 is much less expensive to target and leads to more opportunities. Finally, AI can assist marketing agencies in rebuilding their ads to meet the needs of a more profitable target group.

Creating ads

A marketing agency will create new ads based on the previously identified customer segments. Marketing agencies frequently use responsive advertisements when developing an ad in Google Ads. Responsive advertisements consist of three headings and three descriptions. Based on the search terms the marketing agency previously filled out, it is now possible to decide which headings and descriptions will be presented to the audience using Google Ads' algorithm, which is built on Natural Language Processing (NLP) and Natural Language Generation (NLG). These search terms demonstrate the ad's relevance. Finally, users can now interact with advertisements that are most likely relevant to them via the web method based on the search terms they have entered.

Targeting customers with hyper-personalised ads

It is now possible to predict which products and/or services a customer will most likely purchase next using AI systems and big data. Targeting customers with hyper-personalised ads allow marketers to maximise their advertising budget and customer data. Furthermore, a customer will not have to interact with online advertisements in which they are unlikely to be interested.

Using customer data to create a comprehensive overview of a customer's needs and desires is critical when developing a hyper-personalised ad. Marketing firms explained that targeting men and women separately by including a picture of a man in a male-oriented advertisement and a picture of a woman in a female-oriented advertisement improved ad performance. However, due to intentional discrimination in online advertising, ad platforms have removed the option of targeting users by gender separately.

Analysing ad performance

Marketing agencies notice that by not being involved in the distribution of online ads, AI systems may draw their own (wrong) conclusions. In some cases, this has resulted in an AI system targeting the most profitable target group rather than focusing on the best product and/or service's best target group.

An example is when a company that rents out holiday homes in Ibiza notice that its advertisements initially focus on older people react significantly to young adults. Because there have been numerous negative experiences with renting out holiday homes to young adults, the company has decided to interfere with the AI system and prevent it from targeting young adults despite this being the most profitable target group. The AI system does not consider its context and will, in most cases, focus on the most profitable target group.

Following the intervention, the AI system will reapply the feedback and learn not to target young adults. In short, analysing ad performance and intervening with AI systems improves AI system performance.

Optimising advertising budgets and ad performance through auction systems

As previously stated, analysing the ad's performance allows the marketing agency's audience to know what works and what does not. These lessons can create new advertisements for (new) customer segments.

Ad platforms use AI for auction strategies. Marketing firms use these auction strategies to launch advertisements for their clients. This is accomplished by informing the platforms that, rather than focusing on the click price (the amount a marketing agency must pay whenever a user clicks on their advertisement), they will focus on selling as many clicks as possible for a large sum.

By bidding for value, marketing agencies can tell the ad platform that the sale will cost them a maximum of 5 euros. If marketing agencies have to pay 10 or 15 cents per click in that case, it is no longer necessary. By the end of the day, 5 euros have been spent. In this case, the ad platform's algorithm will ensure that a match can be made with the appropriate target group.

Based on signals, such as: How many people will interact with this ad? How many competitors are there in the neighbourhood? How many units of this product were sold last year, and so on. An AI system can determine the cost-per-click of the ad. Typically, determining how much advertisers pay for a click price can take up to a week of work. As a result, the marketing agency would frequently have to pay more or less for the clicking price, impacting performance. Finally, marketing firms that use AI systems in their auction strategies can save time and reinvest it in strategy or customer service.

3.2.2. Strengths, Weaknesses, Opportunities, and Threats in the application of Artificial Intelligence in Online Advertising

Through the interviews, this thesis has identified the following Strengths, Weaknesses, Opportunities and Threats of AI in online advertising.

Table 10

Strengths, Weaknesses, Opportunities, and Threats of AI in Online Advertising (Own elaboration)

Codes	Description	Interviews
Strengths		
Optimisation	AI in online advertising and – marketing allows one to organise, evaluate, process, and interpret significant amounts of data. AI finds linkages between patterns that humans would not be able to identify. In addition, AI contributes to new technological developments.	*"With the application of AI in technological development, the sky is the limit"* – Interviewee 18
Efficiency	Even if a marketing agency has brilliant people working on a project, employees will still get tired and might make mistakes, leading to biases. AI does not get tired and can be employed for repetitive tasks, increasing processes' efficiency.	*"The opportunities in Artificial Intelligence lie in its supportive elements. It allows teachers or marketers to save time to invest in their students or clients. In contrary to humans, it cannot get tired."* – Interviewee 9
Weaknesses		
Black Box	AI systems can consist of a so-called black box, which means that sometimes even the developers of the AI system do not know why certain decisions are being made. This can be a weakness, especially if someone would like to know why they are targeted for certain advertisements and why an AI system has made certain decisions.	*"The biggest interest of Artificial Intelligence is that we do not understand it. Due to the black box of AI systems, marketers do not understand the systems they are working with eve. The users that interact with the output of the AI system also do not understand it or sometimes do not even know that they are being subjected to an AI system."* – Interviewee 2

Unintended Outcomes	Unintended outcomes are outcomes that should have been designed, planned, or implemented to happen in an AI system. An example of an unintended outcome is the creation of responsive ads on logical texts. Based on a Google Display Ad, this would lead to the following result: Bags to buy – Buy your bags here – Buy the best type of bags here Even though AI systems are quite capable of interpreting and understanding human language, the systems can still forget what logical texts look like. The example above might look to an AI system as explicit. However, a human reading the ad would most likely be disturbed by the repetition of keywords and choose to avoid interacting with it, missing the entire goal. Interviewed marketing agencies stated currently not following any procedure and/or laid out steps when unintended outcomes have been detected within their online advertising campaigns. Marketing agencies try to limit these risks by limiting the scope in which the AI system is allowed to operate. Or to prevent unintended outcomes by keeping the human in the loop.	*"Unintended outcomes are where decisions or predictions happen that were not intended by ever designed, planned, or implemented in AI."* – Interviewee 8
Continuity & Scalability	Continuity & scalability: Continuity and scalability are defined as strengths and weaknesses. AI system does what it is programmed to do. In most cases, this will be performing repetitive tasks and delivering better quality at a higher speed and scale than humans can do. However, there are biases within the AI system. In that case, these biases will be repeated at a much larger scale and have	*"Scalability is a big problem of digitisation. If one worker performs one wrongful task, that is not much of an issue. However, as soon as you enter the algorithm and systematise that into all situations, you will have larger groups and a more significant impact.*

	more significant outcomes than if a human performs the same job.	*For example: If you have 100 applicants and your personal bias means that people with a black skin colour have a 1% less chance for a job, your algorithm will learn your personal preference. And if we employ this algorithm for over 1000 applicants, then we have a completely different problem at once."* – Interviewee 3
		But also: *"Within the advertising field, we have noticed the high level of continuity of AI. AI 'continues' to workday and night. Even if real-life employees carried out these same tasks, we would still need five full-timers to maintain the same level of efficiency and quality."*– Interviewee 15
Opportunities		
Growth Opportunities	Applying AI within daily practises saves companies time which can be reinvested into other practices leading to growth opportunities and more innovation.	*"Initially, we were scared that applying AI within our daily practices could lead to unemployment. However, we save time by spending less on practical tasks, such as increasing or decreasing a click price. We can now spend this time on strategy, creativity and contact with the customer so that the customer enjoys a better result and experience. Applying AI within our business practices is not a threat but an opportunity. Thanks to new and upcoming AI applications, we have scaled up the number of hours for customers. And our company keeps growing."* – Interviewee 5

| Availability in Big Data | AI systems run on data. The higher the quality of the data, the better the outcomes. Making correct decisions and creating hyper-personalised ads requires much qualitative data. Luckily for marketing agencies, recent technological developments are revolutionising the targeting of customers. Information and Communication Technologies (ICT) can interpret large datasets with detailed information about individual consumers' characteristics and preferences through AI. In addition, these datasets are increasingly becoming increasingly available because more economical (webshops) and social (Social Media) occur online. | *"This is what we are very strongly trying to proclaim within marketing. Ensure that you have your data centre to link your external sources, such as your land register data and internal sources, your CRM system, your webshop and online channels. You can then evaluate this and feed this information back to your online channels. Based on this data, you can target your customers in a much more targeted way."* – Interviewee 17

"In the example of medical science, this is very interesting. There is a mountain of data available. If you can, say, make new diagnoses or develop new treatment methods that you had never thought of before, then yes, that is very interesting." – Interviewee 10 |

Threats

Upcoming Regulations	Most marketing agencies agree with regulations regarding the application of AI within their business practices and data processing as long as it does not comply with their business model.	*"As a company, you must deal with laws and regulations, but we are also marketers with a business model. If we are not allowed to gather big data to target customers and perform our business practises, we can no longer exist."* – Interviewee 5 *"I agree with regulating AI so that it does not lead to unintended harmful outcomes. On the other hand, I also find it very difficult in a world where we all sell something, from a company point of view, that the world is becoming so closed on a privacy level. That is a disadvantage, not because I just want to be able to spam and harass everyone unlimited. But I do notice that today it is and is becoming extremely difficult to contact people."* – Interviewee 14
Polarisation	AI systems employ feedback loops to implement outcomes as input data. This input data could consist of the content a web user enters online, such as recommendations sent out by recommender systems, relationship ties on social networks or hyperlinks on web pages. By reusing the same content (output) that a user interacts with as input data, biases could be induced, leading the users to be trapped in a polarised bubble with no access to other content. Diverse content and diverse opinions lead to better availability of information. However, continuously interacting with the same content could lead to polarisation. Polarisation's danger lies in the possibility of it leading to harmful conspiracy theories.	*"People are put into 'content boxes' by AI to see a certain type of content on the internet and do not interact with others. This pushes them into a certain direction of thinking without also weighing in others because they do not get to interact with them."* – Interviewee 1

3.2.3. Causes Unintended Outcomes Artificial Intelligence

This Chapter will examine why and how unintended outcomes occur in online advertising due to the application of AI and how unintended outcomes can be prevented. From the findings of the interviews, this thesis has identified that there are several causes for unintended outcomes:

Business Understanding

Unintended outcomes occur when a company fails to define a clear problem definition for the AI system to solve. If a company is uncertain about what to do with an AI system, it should avoid implementing one.

"Not every company should employ AI within their business practices. If you do not have a problem or challenge that you would like AI to solve, do not employ it." – Interviewee 6

Data Understanding

The data quality must be verified before it can be used in an AI system. If this does not occur (correctly), the dataset may still contain prejudices or other types of biases from the developer.

"It is not a secret that humans are biased, that many institutional agencies are virtually biased and most often are biased in the same direction. They are racist; they are sexist; they are very, very disciplinary. They are sometimes blind to the specificities of people. It is hard to claim this would not have happened with AI." – Interviewee 2

Data Preparation

The rule favours quality over quantity regarding data in AI systems. Unnecessary information about a target group should be removed (cleaned) to avoid unintended outcomes from the AI system.

"I would say that in 90% of the cases, unintended outcomes occur because of the data. We employ ever larger data sets to train ever larger models. This brings us to a point where it is almost impossible to check the data to ensure that there are no harmful sentences or images or whatever in there." – Interviewee 8

Modelling

Algorithms of various types can be used in ML Models. As previously defined, ML Models can be class algorithms (see of algorithms see Table 3. When the incorrect algorithm is used for inappropriate activity, unintended outcomes occur.

"Specifically, neural networks also have the reputation of being a black box, which means that outcomes that are not wanted can happen that can not be clarified." – Interviewee 17

Evaluation

The AI system determines whether the input results in the expected output. When this does not occur (correctly), the biases that are still accidentally implemented within the system c The model and output can still be adjusted by evaluating the data still be adjusted to avoid unintended outcomes.

"When in a testing phase of an AI system the outcomes are discriminative, then you could change either the cohort of users or the testers, for instance, so that you diversify the data." – Interviewee 9

Deployment

Using the output of an AI system or what an AI system is applied for could impact possible unintended outcomes. When the AI system is involved in a different setting than it was trained or built for, it could lead to unintended outcomes. However, it could also be that the output is applied for a different purpose.

"And then it is also how the results are used. How do you handle uncertainty? When the algorithm says I am 60% sure it is A, there is also a 40% chance of B. How do you handle that? Do you say, okay, then I take A? Or do you say, okay, I am not sure?" – Interviewee 8

Inspired by van Giffen (2022), in combination with the conclusions and recommendations from the interviewees, this thesis has designed the following mitigation methods for unintended outcomes:

Table 11

Mitigation Methods Unintended Outcomes in Artificial Intelligence (Own elaboration)

Phase	Method	Action	Quote Interviewees
Business Understanding	A) Increase understanding goal	Companies need a clear business understanding before implementing it within their business practices. What is the goal of implementing AI? Moreover, is AI the right tool for this matter?	*"Companies have a predetermined goal, but how they reach it is not entirely determined. So perhaps in the way of reaching that goal, they end up having these externalities (unintended outcomes)."* – Interviewee 16
	B) Represent all stakeholders	Generate dialogue with a broad set of stakeholders so that the impacts of AI can be brought to the table from diverse perspectives.	*"Generate dialogue with a broad set of stakeholders so that the impacts of AI can be brought to the table from very diverse perspectives."* – Interviewee 9
	C) Discuss technical – and social consequences	Discuss with all the stakeholders what the technical and social consequences will be when this AI system is deployed in real. How will it function in a real-life context? Moreover, what will the impact be if the AI system makes the wrong conclusions?	*"We conduct reviews on products, systems, and innovations. New services go through an entire review by people who are therefore specialised and think about what the consequences are."* – Interviewee 18
Data Understanding and – Preparation	D) Include experts on data selection	Include AI experts that can help identify features that should be included for training the model. For instance, have AI experts for the correct labelling of the training data to prevent biases.	*"What we want in the marketing field is to retrieve the data as much as possible to your infrastructure, sources, or systems. Suppose you are not an expert in marketing, and you have campaigns running. Try to collect as much as data so that it is possible*

			to learn from this data and that data agencies can you help substructures data." – Interviewee 17
	E) Targeted data augmentation	Improve the balance of the dataset by equalising overrepresented – and underrepresented groups in a dataset. If the current dataset contains 80% men and 20% women, this must be equalised to create fair outcomes.	*"The first step is that I believe most of them (unintended outcomes) comes from the data to check your data better to ensure it comes from what it contains. And then I think also the explainability part, the transparency part. Being able to understand what your algorithm does can help mitigate unintended outcomes."* – Interviewee 12
	F) Adopt 'Transparency by design principles	Often algorithmic biases are inherited because the input data is not representative of the sample, leading to biases toward specific categories of people. Fairness is ensured by providing transparency over how data is collected in decision-making.	*"Companies should develop an annual report on transparency in which they illustrate how, why and for what reasons they have applied certain data and AI systems."* – Interviewee 10
Modelling	G) Stimulate explainable AI	Increase the transparency of algorithmic systems. AI systems are based on deep-learning techniques in which the intermediate layers between the input data and the algorithmic output are often defined as a "black box". Explainable AI can substantially contribute to understanding how these automated systems work.	*"AI systems are so complex, and at some point, the systems designers do not know what will come out at the other end. So that is a huge challenge for a legislator to address. And the companies themselves, because they say, well, how can we provide explainable AI if we do not know ourselves?"* - Interviewee 16
	H) Keep the human in the loop	Impose human oversight to be critical of algorithmic recommendations, especially	*"The weakness of the system that we should take into account is that there should still be a human in*

		in high-risk applications, during the development of the AI model, and monitor it when deployed. Humans taking part in the algorithmic decision loop can prevent discrimination from occurring and directly impose that the outcome should be non-discriminatory.	*the loop, that we should not just blindly trust this AI-based system."* – Interviewee 9
	I) Use multiple models	Use multiple models for different groups. For instance, for gender: employ one model for male users and one for female users. This way, 'gender' can no longer influence the outcome of the AI system as both male – and female users are equally targeted when the models are combined into one.	*"Companies should start by implementing Artificial Intelligence within their business practices in small steps, starting with linear regression and/or a simple decision tree. This will help to understand how AI works and how it can work for them and also prevents large biases and unintended outcomes."* – Interviewee 17
Evaluation	J) Technological assessments	Integrate technological assessments that allow marketers to test if their AI systems lead to the right outcome. Compare the AI system with the outcomes that a human would create and compare these outcomes with huge differences.	*"We try to prevent this (unintended outcomes) by testing/experimenting with as many of these algorithms as possible. If you do not experiment, you cannot rule out either. If an outcome does not turn out as expected, it is still a good result because you can include it in your adjustments. As a result, you bring more focus to your current strategy. This piece does not work for these customers in this setting. This makes it easier to reform your algorithm."* – Interviewee 5
Deployment	K) Monitor plan	Develop a monitoring plan for the AI system's deployment in real life. Who will maintain the AI system when the	*"Especially from a technical perspective, many companies employ services with even less transparency today. Even if I (an AI expert) ran it on my machine, it is a black box. And sometimes,*

	audience and/or the system's context changes?	*this AI system can change without me noticing. From one day to another, sending the same document through the same service can produce a different result. This means if I employ a certain service, things can change without me even noticing it if I do not monitor it (the input and output) constantly."* – Interviewee 8
L) Regulate the application of AI	Authorities, big professional companies, NGOs, and, most importantly, SMEs should work together to develop a code of ethics rules that will lead to trustworthy AI systems by design. Regulation positively impacts AI by creating incentives to advance tools to create AI that is not biased and prevents discrimination.	*"Regulation positively impacts AI by creating incentives to advance tools to create AI that is not biased and prevents discrimination."* – Interviewee 12

Concluding from the biases identified within the CRISP-DM Model and these mitigation methods, Figure 4 presents a newly designed CRISP-DM Model:
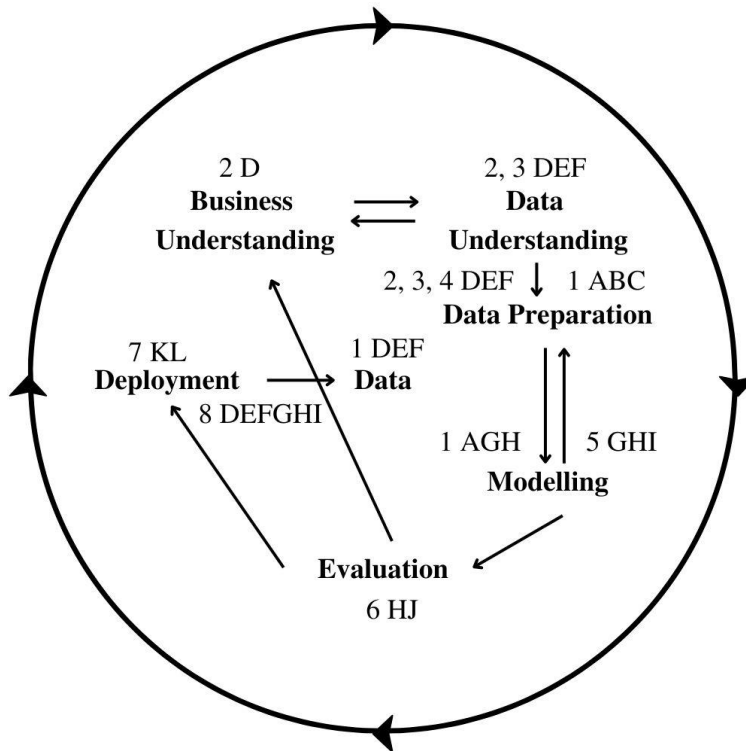

Figure 4

CRISP-DM Model Biases and Mitigation Methods (Own elaboration)

3.2.4. Marketing Agencies' capabilities meeting compliance on the legal framework(s) in the EU

After reviewing how companies employ AI in their online advertising activities and how this may result in unintended outcomes, it is useful to outline how companies meet the legal framework(s) applicable to AI in the EU. The following arguments for and against stronger AI regulations were identified based on the interviews.

Table 12

Interviewee's perspective on AI Act (Own elaboration)

Pro AI Regulation	Con AI Regulation
"Yes, of course, I am in favour of more regulation because I think it is important that advertising is honest and is not confusing." – Interviewee 4	*"The guidelines drawn up by a group of experts within the AI Act are great but translating these guidelines into the practice of companies or legislation and being able to enforce them is another story."* – Interviewee 16
"Technological innovation needs to find a place and needs to be fostered because it is from the perspective of economic growth important." – Interviewee 12	*"If SMEs are not involved in developing the AI Act, the AI Act will not fit them, and it will not be realistic for them to implement. This was also the case for the GDPR, and it was and still is hard for SMEs to comply with the GDPR."* – Interviewee 2
"It is good to look at the risk of systems and determine, based on a risk level, which regulations it must comply with, what condition it is, and that some systems simply should not be used." – Interviewee 10	*"AI – and legal experts have proven that AI, Big Data, and ICT are the three interrelated units. You cannot regulate one thing and expect it to solve the problem. You will have to devise/draw up separate laws and regulations for each sector."* – Interviewee 11
I also noticed many aspects where I would rather be stricter beforehand than afterwards when something happened. It is better to have a bit stricter rules early, and nothing happens than to be more lenient and say: "Yeah, we will be fine." – Interviewee 9	*"The question of whether product safety is the most logical choice for regulating AI. AI has the characteristic that it is changeable and continues to develop during use. Would not food safety legislation be a better source of inspiration then?"* – Interviewee 1

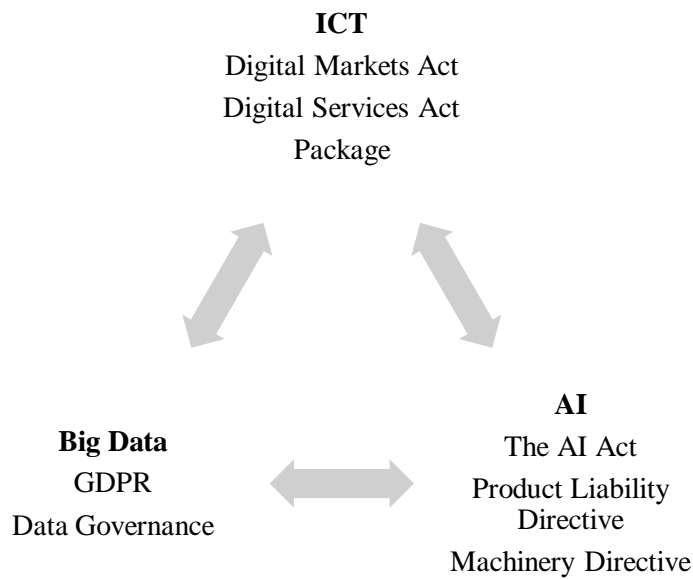The most critical outtakes from the interviewees are:

1. SMEs should be more involved in developing the AI Act as it is challenging to implement the new regulation, as was with the GDPR.

SMEs need help translating the guidelines from the AI Act into practice. Companies applying AI within their advertising activities only comply with the EU's GDPR as a legal framework. As mentioned in Figure 3, marketing agencies apply AI to win various activities. This means the new AI Act might also apply to their activities. Companies applying AI were unfamiliar or not current with the proposed AI Act. They interpreted new regulations to threaten their business model. SMEs fear that when not involved in developing new laws and regulations (as has happened with the AI Act and GDPR), they will find the legislation unfit for them, and it will not be realistic for them to implement it. As a result, SMEs should be included in developing new laws and regulations that they are most likely to comply with. This would enable them to demonstrate the various activities in which they employ AI and its impact on these.

As featured in Figure 2, the AI Act requires conformity assessments or setting up quality management for companies to control their (high-risk) AI systems. However, SMEs need to be able to afford these assessments or the establishment of AI-Risk Management Programs, as SMEs often exist out of only two or three people. These measures require a considerable amount of money and time. In addition, it also takes knowledge either in-house (which SMEs do not frequently have) or by hiring experts with money that SMEs also do not have. Therefore, companies ask in most legislations for SME exemptions, – support or – exceptions.

2. AI, ICT, and Big Data should also be regulated to prevent harm to technological systems.

Figure 5 strengthens the theory that ICT, Big Data and AI are a holy trio correlated with each other. This figure was established by combining the laws and regulations in Table 7. Each sector is represented by different proposed or already adopted laws. In conclusion, each industry and law must also identify interevent unintended outcomes.

ICT
Digital Markets Act
Digital Services Act
Package

Big Data
GDPR
Data Governance

AI
The AI Act
Product Liability
Directive
Machinery Directive

Figure 5

Correlation of AI, Big Data, and ICT (Own elaboration)

3. The AI Act should not include the Product Liability Directive but should focus on the Food Safety Legislation Act to improve fair and safe AI systems.

Compared to the systems and products that the Product Safety Directive assesses, the AI systems need to align. The AI Act has been developed based on the Product Safety Directive and the Machinery Regulation Directive. Both were established for the first time 20 years ago. Based on both acts, the EC decides whether machinery products, such as washing machines and microwaves, are good enough and are not harmful to people.

Nevertheless, a drawback of AI systems is that they can operate as a black box and lack transparency, as was previously mentioned in the SWOT Analysis. This implies that the operation and output of AI systems are only sometimes understood. Compared to microwaves and washing machines, whose functions are far more visible.

AI has the characteristic that it is changeable and continues to develop during use. According to the point of improvement of the Future Society, it should remain responsive to technological trends and is now incapable of doing so. For these reasons, the Food Safety Legislation is, according to interviewees, a better alternative as this legislation contains more flexibility and required check-ups.

4. A CHECKLIST FOR COMPANIES APPLYING ARTIFICIAL INTELLIGENCE IN ONLINE ADVERTISING

Before the proposal of the AI Act, the EC had the European Ethics Guidelines for Trustworthy AI. This guideline describes the seven critical requirements for Trustworthy AI: Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-Discrimination and Fairness, Societal and Environmental Well Being and Accountability (European Union, 2019).

Additionally, commissioned by the Ministry of the Interior and Kingdom Relations, the Utrecht Data School department at the University of Utrecht developed the FRAIA regulation, an abbreviation for the Fundamental Rights and Algorithm Impact Assessment. The assessment of algorithms enables companies to evaluate if algorithms should be used, what algorithms should be applied for, and whether doing so will not lead to discrimination (Utrecht Data School, 2021).

This thesis summarises its findings in combination with the European Ethics Guidelines for Trustworthy AI and the guidelines from the FRAIA regulation. In addition, these findings are supported by the feedback on trustworthy AI systems from Professor Sennay Gehreab, founder and Scientific Director of the Dutch CIVIC AI Lab. These findings create a checklist which companies can employ to prevent unintended outcomes. According to Professor Sennay Gehreab, four questions are necessary to make algorithms fair and inclusive. In the National AI Course on Ethics, 2022, the professor describes the four questions that need to be asked to develop an algorithm:

1. *"Why: Why do we develop algorithms?*
1. *Who: Who makes algorithms?*
2. *What: What data do you use?*
3. *How: How do you make algorithms?"* – Sennay Gehreab (Innovation Center for Artificial Intelligence, 2018)

The four steps have been divided into the phases – and mitigation methods of the process of an AI system based on the CRISP-DM Model in Table 11 in Artificial Intelligence. Finally, check questions have been developed that companies can employ to ensure these mitigation methods are applied and untended outcomes are avoided.

Table 13

Checklist Mitigation Methods Unintended Outcomes (Own elaboration)

	Phase	Method	Action	Check Questions
Why	Business Understanding	Increase understanding goal	Companies need a clear business understanding before implementing it within their business practices.	What problem would you like to solve by applying AI? What is the reason for applying an AI system? What is the goal to be achieved by applying the AI system? What is the main goal, and what is the subgoal? Is an AI system the right choice to solve this issue? Which other tools are available to solve this issue?
		Represent all stakeholders	Generate dialogue with a broad set of stakeholders so that the impacts of AI can be brought to the table from diverse perspectives.	Which parties and persons are involved in the development/use/evaluation/maintenance? Who is going to be the end user of this AI system?
		Discuss technical – and social consequences.	Discuss with all the stakeholders what the technical and social consequences will be when this AI system is deployed in real. How will it function in a real-life context? Moreover, what will the impact be if the AI system makes the wrong conclusions?	How will it function in a real-life context? What will the technical impact be if the AI system makes the wrong conclusions? What will the social impact be if the AI system makes the wrong conclusions? What will be the effects of the AI system? How do the effects of the AI system contribute to the problem it is supposed to solve? How is the context in which the AI system operates taken into account? What risks are there for adverse unintended outcomes for end users? How will these risks be reduced and prevented?

What	Include experts on data selection.	Include AI experts that can help identify features that should be included for training the model. For instance, have AI experts for the correct labelling of the training data to prevent biases.	Is the AI system developed internally or by an external party? What are the agreements about the ownership and management of the AI system?
	Targeted data augmentation	Improve the balance of the dataset by equalising overrepresented – and underrepresented groups in a dataset. If the current dataset contains 80% men and 20% women, this dataset must be balanced in order to create fair outcomes.	What type of data will be applied as input data? Where doe this input data come from? How reliable and how high is the quality of this input data?
	Adopt 'Transparency by design principles in Data Understanding – and Processing	Often AI systemic biases are inherited because the input data is not representative of the sample and leads to biases toward specific categories of people. Fairness is ensured by providing transparency over how data is collected in decision-making.	What assumptions and biases lie in the data? (Fairness) What is the influence of these assumptions and biases on the output? Is the data representative of the context in which the AI system will be applied? Is it transparent what an AI system does, how it does it and based on which data the AI system creates its output?
	Ensure safe archiving of data	Personal data databases fall under the General Data Protection Regulation (GDPR) and the Archives Act. Both are important when storing or destroying documents containing personal data.	How does the AI system protect personal data relating to individuals in line with the GDPR? How did the AI system perform on the Data Protection Impact Assessment (DPIA)? How and where are the input – and output data archived? Whom is the input data shared with? Whom is the output data shared with? How is access to the data controlled?

Who	Modelling	Keep the Human in the Loop	Impose human oversight to be critical of AI systemic recommendations, especially in high-risk applications, during the development of the AI and monitor it when it has been deployed. By humans taking part in the AI systemic decision loop, companies can prevent discrimination from occurring and directly impose that the outcome should be non-discriminatory.	Does the AI system create a similar output as a human would create? Why does or does not the AI system create a similar output as a human would create? Have the humans in the loop been trained on imposing oversight?
How	Modelling	Stimulate explainable AI	Increase the transparency of AI systemic systems. AI systems are based on deep-learning techniques in which the intermediate layers between the input data and the AI systemic output are often identified as a "black box". Explainable AI can substantially contribute to understanding how these automated systems work.	Is it transparent what an AI system does, how it does it and based on which data the AI system creates its output? (Transparency) For whom (which stakeholders) will the AI system be made transparent? How is the AI system made transparent? Can the choices of an AI system be clarified to someone who does not understand it?
		Use multiple models	Use multiple models for different groups. For instance, for gender: employ one model for male users and one for female users. This way, 'gender' can no longer influence the outcome of the AI system as both male – and female users are equally targeted when the models are combined into one.	What type of AI system is employed? a. A non-self-learning AI system in which one specifies the rules that the computer must follow. (supervised learning) b. b. A self-learning AI system in which the machine learns about the data's patterns (machine learning). (unsupervised learning) Why is this type of AI system chosen?

				Why is this AI system best suited to achieve the intended objectives? What are alternatives available? Moreover, why were these alternatives not chosen?
	Evaluation	Technological assessments	Integrate technological assessments that allow marketers to test if their AI systems lead to the right outcome. However, compare the AI system with the outcomes that a human would create and compare these outcomes with huge differences.	How are the results of the AI system evaluated? (Accuracy) What is the accuracy of the AI system? Based on which evaluation criteria has this accuracy been determined? What are the measures/procedures when the AI system does not meet the evaluation criteria? How often/how much is the AI system wrong?
For what?	Deployment	Monitor plan	Develop a monitoring plan for the AI system's deployment in real life.	What happens to the results (output) of the AI system? Which decisions are based on the outcome of the AI system? Who will maintain the AI system when the audience and the system's context change? How is it possible for people to monitor the quality of the AI system? Is there enough qualified staff now and in the future to manage, review, and adjust the quality of the AI system?
		Regulate the application of AI.	Authorities, big professional companies, NGOs, and, most importantly, SMEs should work together to develop a code of ethics rules that will lead to	Through which regulations will AI systems be developed? How are the key stakeholders involved in the development of the AI system?

trustworthy AI systems by design. Regulation positively impacts AI by creating incentives to advance tools to create AI that is not biased and prevents discrimination.	How is it ensured that the development of AI systems meets the most critical requirements of regulations, such as the AI Act?

5. DISCUSSION

This Chapter starts by interpreting the results of this thesis and discussing their implications. Finally, the limitations of this thesis are outlined.

5.1. Interpretation of the Results

The results of this thesis indicate that companies are increasingly dependent on the application of AI within their online advertising activities. From the findings of the interviews, it can be deduced that when all competitors employ AI, an individual company also tends to employ AI to remain relevant for its customers. Additionally, the results of this thesis illustrate that unintended outcomes do happen, but not as often or as drastic in online advertising, as illustrated in the theoretical research of Bandy (2021).

Based on the findings of the interviews, it is possible to conclude that the majority of marketing agencies are aware of the potential harms of AI. However, unintended outcomes are avoided by keeping the human in the loop, focusing on a small scope in which an AI system can operate independently, and evaluating the outcomes of the AI system. Additionally, an example of the Case Study demonstrated that unintended outcomes within online advertising happen regardless of the advertiser's actions.

In order to prevent harmful unintended outcomes from happening, the EC has proposed the AI Act. The results from the Legal chapter demonstrate that the opinions on adopting this new law are divided. Those favouring the AI Act argue that the new law will stimulate transparent advertising. Those against it argue that it may harm marketing agencies' business models. Additionally, those against argue that, in particular, SMEs are incapable of implementing the AI Act standards. For example, they mentioned the difficulties for SMEs in complying with the GDPR.

The following paragraphs focus more on potential explanations why the results from this thesis did not meet the literature.

Transparency interviewees

A question included within the interviews focused on the occurrence of harmful unintended outcomes within marketing agencies. According to the literature, these unintended outcomes are widely documented (European Digital Rights, 2021; O'Neil, 2017).

However, based on the findings of the interviewees, it can be concluded that these unintended outcomes within online advertising did not happen as often as one might think after

reading about them in the literature of the European Digital Rights organisation (2021) or O'Neil (2017).

It could be perceived that a marketing agency creating unintended outcomes as a consequence of applying Ai might not be transparent about this. The occurrence of unintended outcomes due to the application of AI could be considered harmful to the brand image of a marketing agency. Thus, a marketing agency may not be as transparent about these to prevent 'bad publicity'.

Scalability

One of the methods to prevent unintended outcomes from happening within online advertising is only to let the AI system operate in a minimal scope. Within this small scope, potential damages that could occur are small and easy to identify and mitigate.

Suppose the AI system was allowed to operate in an enormous scope. In that case, the damages of a malfunctioning AI system may be more significant, as Ali et al. (2019) and Lambrecht & Tucker (2019) demonstrated in their research. Entering a biased algorithm into a larger scope leads to systematising its deviated outcomes.

Marketing agencies retain the small scope in which the AI system is allowed to operate. Additionally, they include different methods, such as keeping the human in the loop, that prevent these unintended outcomes from happening.

Business Model

Akter et al. (2022) argue that companies intentionally manipulate AI systems to create unintended outcomes to create an unfair advantage within marketing and increase their sales revenues. This thesis has not found evidence that supports this statement.

On the contrary, based on the findings from the interviews, this thesis found more evidence of marketing agencies preventing unintended outcomes within online advertising. Furthermore, marketing agencies creating advertising campaigns for their clients will ensure that the AI systems work as expected. Malfunctioning AI systems would produce incorrect results, resulting in lower customer satisfaction and turnover. Finally, avoiding unintended outcomes is in a company's best interest.

5.2. Implications

The results of this thesis are in some respects consistent with previously administered research by Ali et al. (2019), Blass (2019) and Lambrecht & Tucker (2019).

Literature suggests that unintended outcomes within online advertisements could harm users (Akter et al., 2022b). However, from interviews, it can be concluded that marketing agencies implement mitigation methods that prevent these potentially harmful unintended outcomes. Therefore, the theory-described unintended outcomes have not been experienced by marketing companies applying AI within their online advertising activities. That being said, this thesis conducted interviews with 18 participants with a non-participating ratio of 74.3%. It cannot be said that because the interviewees did not experience harmful unintended outcomes, it did or does not happen in online advertising.

Additionally, the same literature suggests that the harmful unintended outcomes within online advertisements are not the intention of an advertising platform or an advertiser (Ali et al., 2019; Blass, 2019; Lambrecht & Tucker, 2019). Interviewees support this statement and argue that unintended outcomes could occur through insufficient business understanding, biased datasets or faulty algorithms. Another cause for unintended outcomes is mentioned by Ali et al. (2019), that argue that algorithms can be a 'black box'. Advertisers may not always understand the features that the AI model uses to create an outcome and may, therefore, be unaware of the causes of potential discriminatory decisions. Interviewees stated that by evaluating the outcomes of the AI system, they were able to prevent harmful unintended outcomes from happening.

Despite the fact that interviewees stated that they were capable of preventing unintended outcomes from happening, the EC has proposed the AI Act. The objective of the AI Act is to prevent all harmful outcomes caused by AI from happening. Based on the literature, the conclusion could be drawn that the AI Act contained loopholes and exceptions (Bringas Colmenarejo et al., 2022; Veale & Borgesius, 2021). The opinions on the adoption of the AI Act among interviewees were divided. Marketing agencies stated that the AI Act would damage their business model. Especially given that they have also encountered this during the implementation of the GDPR and the exclusion of SMEs from the AI Act's development. On the contrary, legal experts favoured establishing the AI Act. In their opinion, the AI Act would prosper honest and transparent advertising.

These results matter because they provide evidence that companies have taken methods to prevent harmful unintended outcomes. The interview findings prove that these methods are, to some extent, effective. This thesis's results also suggest that the proposed AI Act may be too

strict for SMEs. Based on this conclusion, it would be better if the checklist were presented as a regulation for preventing AI systems that could lead to unintended outcomes.

5.3. Limitations

The input from 18 interviews underlines the findings of this thesis. More interviews could give more insight into how companies employ AI within their advertising activities, how companies prevent unintended outcomes, and what the opinion is on stricter AI regulations and legislation.

This study included the opinion of 18 interviewees out of the 70 potential interviewees that were contacted. The lack of participation of interviewees has led to a non-participation ratio of 74.3%. Therefore, the findings cannot indicate if other Marketing - AI – and/or Legal Experts would have shared a different perspective on the subject.

In addition, other studies have examined the unintended outcomes of online advertising by creating and running advertisements themselves, applying Facebook as an ad platform (Ali et al., 2019; Lambrecht & Tucker, 2019). This thesis could not practise this method due to time and cost constraints and therefore drew on the results from the literature and interviews to support its conclusions. Not checking if unintended outcomes occur in self-designed ads to measure potential unintended outcomes could be considered a limitation.

6. SUMMARY AND CONCLUSIONS

The final Chapter of this thesis summarises the main results and, by doing so, answers its main research questions. Furthermore, it provides recommendations for companies applying Artificial Intelligence (AI) within their online advertising campaigns. Finally, this Chapter features potential future search directions.

6.1. Conclusions

This thesis investigates how the application of AI within online advertising campaigns can lead to unintended outcomes and with which legal framework(s) companies should comply to mitigate these measures. This thesis explores which mitigation methods companies can administer against unintended outcomes caused by online advertising due to the application of AI.

This research has given insights into how AI is being applied within online advertising by companies, how AI can lead to unintended outcomes and how mitigation methods can prevent them from happening. Subsequently, the thesis reviews the dynamic legal landscape in which companies operate by answering the research questions. The conclusion is formulated after answering the following research questions:

How are companies currently implementing Artificial Intelligence within their advertising campaigns?

Companies employ AI for the following activities: 1. Identifying new customer segments, 2. Creating ads applying AI-enabled NLP (Natural Language Processing) and NLG (Natural Language Generation) tools, 3. Target customers with hyper-personalised ads applying data that products might like and/or will likely buy next, 4. Evaluate the performance of ads and, if necessary, intervene with the AI system to improve the performance of new ads, 5. employ AI-enabled auction systems to regain focus on strategies and customer services and increase growth opportunities.

To what extent does employing artificial intelligence lead to unintended outcomes in online advertising?

With an AI system, the following biases can occur: Social –, Measurement –, Representation –, Label –, Algorithmic –, Evaluation –, Deployment – and Feedback Bias. These biases can be caused by: Not having a clear problem definition of what the AI system needs to solve; Low quality of data due to not properly cleaning the dataset or including prejudices; Applying the inappropriate type of algorithm to an activity; Not evaluating the AI system; Applying the AI

system in a different setting than it was trained or built for. The following mitigation methods have been identified to prevent these biases from happening. Per step of the CRISP-DM Model, mitigation methods to avoid unintended outcomes from happening have been identified. The Business Understanding phase represents the steps: Increasing goal understanding, representing all stakeholders and discussing technical – and social outcomes when applying an AI system. The Data Understanding and – Preparation phase includes experts on data selection, improving the dataset through targeted data augmentation, and adopting 'Transparency by design principles. The Modelling stage represents: Stimulating explainable AI, keeping the human in the loop and applying multiple AI models. Through technological assessments within the Evaluation-phase, evaluation biases can be detected and prevented. Finally, biases can be contained within the Deployment-phase by monitoring the AI system and regulating the application of AI.

How are companies employing Artificial Intelligence for their advertising campaigns currently meeting the EU's legal framework(s)?

Companies applying AI within their online advertising campaigns only meet the criteria of the GDPR (General Data Protection Law). However, their opinions regarding the implication of the AI Act are primarily negative. The legislation needs to be more relaxed according to them, and rules related to the level of risk per AI system are unrealistic to implement for SMEs (Small and Medium Enterprises) due to costs, time and/or workforce.

For this reason, this thesis provides new points of improvement of the AI Act, which still need to be identified by other studies and which could be included within a new proposed AI Act. The most important conclusions summarised are that marketing agencies have adopted mitigation methods to prevent harmful unintended outcomes from happening within their online advertising activities. In addition, it is expected that simultaneously with the GDPR, SMEs will struggle to comply with the standards of the AI Act.

Better involvement of SMEs in the development of the AI Act

SMEs should have been more involved in the AI Act's development. SMEs are expected to be assisted in implementing the new AI Act's standards (Three figure step). As a result, solutions for SMEs may include exemption, exception, and redemption. Furthermore, SMEs need assurance that the AI Act will not harm their business model to comply with it.

SMEs should have been more involved in the development of the AI Act. It is expected that SMEs will need help to implement the standards with the new AI Act (Three figure step).

Therefore, solutions for SMEs could be exemption, exception and redemption. In addition, reassurance that the AI Act will not threaten SMEs' business model would be necessary for SMEs to comply with the AI Act.

Focus on regulating AI, ICT, and Big Data as a holy trio.

Big Data, ICT, and AI cannot function without each other. Biases within AI systems can occur in various ways. The data is flawed, the companies operate as gatekeepers, or the AI systems are biased. Each of these three components could lead to unintended outcomes. This is why each must be regulated through regulations and legislation. Concluding that the adoption of the AI Act alone is not enough.

Big Data, ICT, and AI are strongly connected. Biases in AI systems can occur in a variety of ways. The data is faulty (Big Data), companies operate as gatekeepers (ICT), or AI systems are biased (AI). Each of these three components may lead to unintended outcomes. As a result, each component must be governed by laws and regulations, such as the Digital Services Act Package, GDPR and AI Act. Conclusion: Adoption of the AI Act alone will not suffice.

Apply the Food Safety Legislation instead of the Product Liability Directive for fair and safe AI systems.

Instead of the Product Liability Directive, the Food Safety Legislation is a better option for controlling the changeable and continuous development that AI typifies. The AI Act remains a flexible and better response to technological trends by implementing Food Safety Legislation rather than the Product Liability Directive.

Finally, the combination of these three research questions resulted in the creation of a checklist. The checklist employs methods for mitigating algorithmic biases. It includes checking questions companies can employ to ensure that these mitigation methods are implemented and that unintended outcomes are avoided.

6.2.   Contributions

This thesis adds to the existing academic literature on artificial intelligence, algorithmic biases, online advertising, and potential legal framework(s) by investigating how companies combat unintended outcomes and how this relates to the recently proposed AI Act. The thesis results collectively create a checklist that marketing agencies can employ to mitigate or prevent biases when applying AI systems.

Furthermore, the checklist enables SMEs to evaluate or recently begun applying AI in their business practices to identify their strengths, weaknesses, opportunities, and threats and how these can be addressed.

6.3.   Recommendations

This thesis can be administered when a company is applying(or is exploring the employ of) AI within their business practices and/or would like to find out how to prevent harmful unintended outcomes from their business practices. Governmental companies could also employ it to identify more improvement points within the AI Act.

Finally, this thesis concludes its recommendations by providing several potential future research directions that could be identified based on its outcome.

The CRISP–DM Model is redesigned in this thesis and now includes mitigation methods to prevent unintended outcomes. However, further research should focus on whether these mitigation methods are applied in online advertising and how these mitigation methods could be improved.

Another research opportunity would be to investigate the real-world applicability of the checklist designed for companies to prevent unintended outcomes. In addition, it would be helpful to research what companies that employ the checklist should do to avoid unintended outcomes.

During the writing of this thesis in March 2023, the AI Act has only been a proposal for now. It would be an interesting future research direction to focus on how the AI Act will develop itself and if it would evaluate the opinions and points of feedback outlined within this thesis. Furthermore, it would be interesting to investigate how/if companies would need to adapt their business model to meet the requirements of the AI Act and what the impact of this might be.

REFERENCES

The E.U. AI Act: A Risk-Based Policy Approach to AI Applications - YouTube. (n.d.). Retrieved November 26, 2022, from https://www.youtube.com/watch?v=OhBsDIG8ioM&t=1543s

Access Now Europe. (2021). *Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act*. https://ec-europa-eu.ezproxy2.utwente.nl/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665462_en

Akter, Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, *144*, 201–216. https://doi.org/10.1016/J.JBUSRES.2022.01.083

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 199. https://doi.org/10.1145/3359301

Andreou, A., Silva, M., Benevenuto, F., Goga, O., Loiseau, P., & Mislove, A. (2019). *Measuring the Facebook Advertising Ecosystem. February*. https://doi.org/10.14722/ndss.2019.23280

Ann, M., Data, B., Neil, C. O., & Lagatta, T. (2022). *Conscientious Classification : A Data Scientist ' s Guide to Discrimination-Aware Classification Authors : June 2017*.

Asplund, J., Eslami, M., Sundaram, H., Sandvig, C., & Karahalios, K. (2020). Auditing race and gender discrimination in online housing markets. *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, *14*, 24–35. https://doi.org/10.1609/ICWSM.V14I1.7276

Bandy, J. (2021). Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–34. https://doi.org/10.1145/3449148

Bashir, M. A., Farooq, U., Shahid, M., Zaffar, M. F., & Wilson, C. (2019). *Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers*. https://doi.org/10.14722/ndss.2019.23392

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data and Society*, *6*(1). https://doi.org/10.1177/2053951718819569

Benjamin, M., Buehler, K., Dooley, R., & Zipparo, P. (2021). What the draft European Union AI regulations mean for business | McKinsey. In *McKinsey&Company* (Issue August). https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-the-draft-european-union-ai-regulations-mean-for-business

Bessen, J., Impink, S. M., & Seamans, R. (2022). The Cost of Ethical AI Development for AI Startups. *AIES 2022 - Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 92–106. https://doi.org/10.1145/3514094.3534195

Blansett, J. (2019). Digital discrimination. *Library Journal*, *133*(13). https://doi.org/10.1093/oso/9780198838494.003.0004

Blass, J. (2019). Algorithmic advertising discrimination. *Northwestern University Law Review*, *114*(2), 415–468.

Borsci, S., Lehtola, V. V., Nex, F., Yang, M. Y., Augustijn, E. W., Bagheriye, L., Brune, C., Kounadi, O., Li, J., Moreira, J., Van Der Nagel, J., Veldkamp, B., Le, D. V., Wang, M., Wijnhoven, F., Wolterink, J. M., & Zurita-Milla, R. (2022). Embedding artificial intelligence in society: looking beyond the EU AI master plan using the culture cycle. *AI and Society*, *0123456789*. https://doi.org/10.1007/s00146-021-01383-x

Bringas Colmenarejo, A., Nannini, L., Rieger, A., Scott, K. M., Zhao, X., Patro, G. K., Kasneci, G., & Kinder-Kurlanda, K. (2022). Fairness in Agreement With European Values. *Oxford*, 107–118. https://doi.org/10.1145/3514094.3534158

Buiten, M. C. (2019, March 1). *Towards intelligent regulation of artificial intelligence*. European Journal of Risk Regulation; Cambridge University Press. https://doi.org/10.1017/err.2019.8

Cabañas, J. G., Cuevas, Á., & Cuevas, R. (2018). Unveiling and quantifying facebook exploitation of sensitive personal data for advertising purposes. *Proceedings of the 27th USENIX Security Symposium*, 479–495. https://www.usenix.org/conference/usenixsecurity18/presentation/cabanas

Cambridge Dictionary. (2022). *BIAS| English meaning - Cambridge Dictionary*. https://dictionary.cambridge.org/dictionary/english/bias

Chhillar, D., & Aguilera, R. V. (2022). An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research. *Business and Society*, *61*(5), 1197–1241. https://doi.org/10.1177/00076503221080959/FORMAT/EPUB

CNN. (2014). *How to get girls into STEM -- The experts speak - CNN*. http://edition.cnn.com/2014/10/27/world/europe/how-to-get-girls/index.html

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011, June 27). *The case study approach*. BMC Medical Research Methodology; BioMed Central. https://doi.org/10.1186/1471-2288-11-100

Dalenberg, D. J. (2018). Preventing discrimination in the automated targeting of job advertisements. *Computer Law and Security Review*, *34*(3), 615–627. https://doi.org/10.1016/j.clsr.2017.11.009

Datta, A., Datta, A., Makagon, J., Mulligan, D. K., & Carl Tschantz, M. (2018). Discrimination in Online Advertising A Multidisciplinary Inquiry. *Proceedings of Machine Learning Research*, *81*, 1–15. https://proceedings.mlr.press/v81/datta18a.html

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

De-Arteaga, M., Feuerriegel, S., & Saar-Tschansky, M. (2022). Algorithmic fairness in business analytics: Directions for research and practice. *Production and Operations Management*, *31*(10), 3749–3770. https://doi.org/10.1111/poms.13839

Dennis A. Gioia, K. C. (1990). Organization: From Substance to Image? *Organization Studies*, *11*(3), 373–394. https://doi.org/10.1177/017084069001100303

Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior: A framework for engaging customers through social media content. *European Journal of Marketing*, *53*(10), 2213–2243. https://doi.org/10.1108/EJM-03-2017-0182

Dworkin, S. L. (2012, September 12). *Sample size policy for qualitative studies using in-depth interviews*. Archives of Sexual Behavior; Springer. https://doi.org/10.1007/s10508-012-0016-6

EUR-Lex. (2012). *EUR-Lex - 12012E/TXT - EN - EUR-Lex*. Official Journal C 326. https://eur-lex-europa-eu.ezproxy2.utwente.nl/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT

European Commission. (2020). White Paper On Artificial Intelligence - A European approach to excellence and trust. In *Journal of Chemical Information and Modeling*. https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission. (2021). 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. *Com(2021)*, *0106*, 1–108.

European Commission. (2022). Questions and Answers: Digital Markets Act: Ensuring fair and open digital markets*. *European Commision*, *April*.

European Commission. (2019). *HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE*. European Commission.

European Commission. (2022). *Digital Services Act – Questions and Answers General information on the Digital Services Act*. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

European Commission. (2021). *Regulation of the European Parliament and of the Council*. European Commission.

European Commission. (2019). *What is an algorithm in AI? | FUTURIUM | European Commission*. European Commission. https://ec.europa.eu/futurium/en/european-ai-alliance/what-algorithm-ai.html

European Commission. (2022). *Data Governance Act explained | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

European Commission. Competition Policy. (2022). *Digital Markets Act (DMA)*. European Commission. https://www.simmons-simmons.com/en/publications/cl5qyuy0o1h0y0a59y4osbiqx/digital-markets-act-is-adopted

European Digital Rights. (2021). How online ads discriminate. Unequal harms of online advertising in Europe. In *European Digital Rights*.

European DIGITAL SME Alliance. (2022). *European Digital SME Alliance*. European Commission. https://ec-europa-eu.ezproxy2.utwente.nl/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665574_en

European Parliament. (2022). *EU Legislation in Progress Ensuring the safety of machines in the digital age Revision of the Machinery Directive*. European Parliament.

European Parliament and Council of the European Union. (2016). *Art. 9 GDPR – Processing of special categories of personal data | General Data Protection Regulation (GDPR)*. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016. https://gdpr-info.eu/art-9-gdpr/

European Union. (2019). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. European Union. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Facebook. (2020). *How Does Facebook Use Machine Learning to Deliver Ads?* Facebook. https://www.facebook.com/business/news/good-questions-real-answers-how-does-facebook-use-machine-learning-to-deliver-ads

Facebook Business. (2015). *Relevance Score for Facebook Ads*. Facebook. https://www.facebook.com/business/news/relevance-score

Future of Life Institute. (2021). *Feedback from: Future of Life Institute (FLI)*. European Commission. https://ec-europa-eu.ezproxy2.utwente.nl/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665546_en

Giffen, B. van, Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, *144*, 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T., & Potapov, D. (2020). Digital Analytics: Modeling for Insights and New Methods. *Journal of Interactive Marketing*, *51*, 26–43. https://doi.org/10.1016/j.intmar.2020.04.003

Hern, A. (2020). Twitter apologises for "racist" image-cropping algorithm. *The Guardian*. https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm

Hoover, L. (2021). Q. vs. Q. S. | G. B. https://www. gcu. edu/blog/doctoral-journey/what-qualitative-vs-quantitative-stud. (2021). *Qualitative vs. Quantitative Study | GCU Blog*. Grand Canyon University. https://www.gcu.edu/blog/doctoral-journey/what-qualitative-vs-quantitative-study

Huang, M. H., & Rust, R. T. (2021). A Framework for Collaborative Artificial Intelligence in Marketing. *Journal of Retailing*, 30–50. https://doi.org/10.1016/j.jretai.2021.03.001

Imana, B., Korolova, A., & Heidemann, J. (2021). Auditing for discrimination in algorithms delivering job ads. *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 3767–3778. https://doi.org/10.1145/3442381.3450077

Innovation Center for Artificial Intelligence. (2018). *Hij is er: de Nationale AI-cursus - ICAI*. Innovation Center for Artificial Intelligence. https://icai.ai/hij-komt-er-de-nationale-ai-cursus/

Kingsley, S., Wang, C., Mikhalenko, A., Sinha, P., & Kulkarni, C. (2020). Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising. *Md4Sg*, *June*, 1–29.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, *10*, 113–174. https://doi.org/10.1093/jla/laz001

Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management*, *90*, 493–507. https://doi.org/10.1016/j.indmarman.2019.05.003

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, *65*(7), 2966–2981. https://doi.org/10.1287/mnsc.2018.3093

Local Government Association. (2018). *General Data Protection Regulation (GDPR) | Local Government Association*. General Data Protection Regulation (GDPR). https://www.local.gov.uk/our-support/guidance-and-resources/general-data-protection-regulation-gdpr

Low, F. S., & Clifford Lee, W. L. (2021). Developing a humanless convenience store with AI System. *Journal of Physics: Conference Series*, *1839*(1), 012002. https://doi.org/10.1088/1742-6596/1839/1/012002

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion.

Technological Forecasting and Social Change, *175*.
https://doi.org/10.1016/j.techfore.2021.121390

McCombes, S. (2022). *What Is a Case Study? | Definition, Examples & Methods*. Scribbr.
https://www.scribbr.com/methodology/case-study/

Meta. (2023). *Informatie over Instagram-advertenties | Meta Helpcentrum voor bedrijven*.
https://www.facebook.com/business/help/793315701035887?id=377090259688020&content_id=4MrMojDhkr4cNlD&ref=sem_smb&utm_term=dsa-1722193855848&gclid=Cj0KCQiA9YugBhCZARIsAACXxeKALHAf_yFPwTlPfqpBkPuNvW7JRnjtxq7UWV6pzqrsHiPJe6QXBDMaAtHuEALw_wcB

Neuwirth, R. J. (2022). The EU artificial intelligence act: Regulating subliminal AI systems.
In *The EU Artificial Intelligence Act: Regulating Subliminal AI Systems*. Taylor and
Francis. https://doi.org/10.4324/9781003319436

Neuwirth, R. J. (2022). Law, artificial intelligence, and synaesthesia. *AI and Society*, *1*, 1–12.
https://doi.org/10.1007/S00146-022-01615-8/METRICS

Ngai, E. W. T., & Wu, Y. (2022). Machine learning in marketing: A literature review,
conceptual framework, and research agenda. *Journal of Business Research*, *145*, 35–48.
https://doi.org/10.1016/j.jbusres.2022.02.049

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social Data: Biases,
Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*(13).
https://doi.org/10.3389/fdata.2019.00013

O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and
Threatens Democracy. In *Penguin Books Ltd* (Vol. 52).

Rijksoverheid. (2021). *Impact Assessment Mensenrechten en Algoritmes*. 1–95.

Roussinov, D., & Zhao, J. L. (2003). Automatic discovery of similarity relationships through
Web mining. *Decision Support Systems*, *35*(1), 149–166. https://doi.org/10.1016/S0167-9236(02)00102-1

Samuely, A. (2017). *Macy's AI-powered shopping assistant underscores mobile's growing
in-store role*. https://www.retaildive.com/ex/mobilecommercedaily/macys-ai-powered-shopping-assistant-underscores-mobiles-starring-role-in-retail

Satariano, A. (2022). *Meta Fined $275 Million for Breaking E.U. Data Privacy Law - The
New York Times*. New York Times.
https://www.nytimes.com/2022/11/28/business/meta-fine-eu-privacy.html

Scribbr. (2018). *Primary vs Secondary Sources | Explained with Easy Examples*. Scribbr.
https://www.scribbr.com/working-with-sources/primary-and-secondary-sources/

Shah, N., Engineer, S., Bhagat, N., Chauhan, H., & Shah, M. (2020). Research Trends on the
Usage of Machine Learning and Artificial Intelligence in Advertising. *Augmented
Human Research*, *5*(1), 1–15. https://doi.org/10.1007/s41133-020-00038-8

Smuha, N. A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3899991

Stanford HAI. (2022). *Home | Stanford HAI*. https://hai.stanford.edu/

Tafesse, W., & Wien, A. (2017). A framework for categorizing social media posts. *Cogent Business and Management*, *4*(1). https://doi.org/10.1080/23311975.2017.1284390

Tay, L., Woo, S. E., Hickman, L., Booth, B. M., & D'Mello, S. (2022). A Conceptual Framework for Investigating and Mitigating Machine-Learning Measurement Bias (MLMB) in Psychological Assessment. *Advances in Methods and Practices in Psychological Science*, *5*(1). https://doi.org/10.1177/25152459211061337

Tegan, G. (2022). *Semi-Structured Interview | Definition, Guide & Examples*. Scribbr. https://www.scribbr.com/methodology/semi-structured-interview/

University of Cambridge. (2021). *Feedback from: University of Cambridge (Leverhulme Centre for the Future of Intelligence and Centre for the Study of Existential Risk)*. University of Cambridge. https://ec-europa-eu.ezproxy2.utwente.nl/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665626_en

Utrecht Data School. (2021). *Utrecht Data School | Impact Assessment Mensenrechten en Algoritmes*. University of Utrecht. https://dataschool.nl/iama/

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*. https://doi.org/10.9785/cri-2021-220402

VEX Robotics. (2017). *Girls into STEM. March 2018*, 9480. http://www.girlsintostem.co.uk/girlsintostem-1

Yadav, S., Pratap Singh, R., Sree, D., & Vidhyavani, A. (n.d.). E-COMMERCE CHATBOT FOR PRICE NEGOTIATION. *Www.Irjmets.Com @International Research Journal of Modernization in Engineering*, *583*. Retrieved April 12, 2022, from www.irjmets.com

Yu, X., Yang, S., & Tian, H. (2020). Analysis and Research on Behavior-Based Price Discrimination on E-Commerce Platform under Big Data. *Proceedings - 2020 International Workshop on Electronic Communication and Artificial Intelligence, IWECAI 2020*, 83–86. https://doi.org/10.1109/IWECAI50956.2020.00023

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, *23*, 100224. https://doi.org/10.1016/j.jii.2021.100224

Žliobaitė, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, *31*(4), 1060–1089. https://doi.org/10.1007/s10618-017-0506-

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. *Council of Europe*. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-

APPENDIX A: LIST OF INTERVIEWEES

Below is a list of all the participants invited to participate in this qualitative research. For (privacy) and safety reasons, these interviewees' names and contact details are not included in this report but are saved in a password-locked Excel file.

Number Interviewee	Expertise
Interviewee 1	Legal Regulations of AI
Interviewee 2	Legal Regulations of AI
Interviewee 3	Legal Regulations of AI
Interviewee 4	Legal Regulation of AI
Interviewee 5	Technological part of AI
Interviewee 6	Technological part of AI
Interviewee 7	Technological part of AI
Interviewee 8	Technological part of AI
Interviewee 9	Technological part of AI
Interviewee 10	Technological part of AI / Legal Regulations of AI
Interviewee 11	Technological part of AI / Legal Regulations of AI
Interviewee 12	Technological part of AI / Legal Regulations of AI
Interviewee 13	Use of AI in Marketing
Interviewee 14	Use of AI in Marketing
Interviewee 15	Use of AI in Marketing
Interviewee 16	Use of AI in Marketing
Interviewee 17	Use of AI in Marketing
Interviewee 18	Use of AI in Marketing

APPENDIX B: INTERVIEW QUESTIONS

This Appendix includes the interview questions that were shared with the interviewees.

Interview Structure, Question 1 (Marketing)

1. **How are companies currently implementing Artificial Intelligence within their advertising campaigns?**

1. What is your name?

2. What is the name of the organization for which you are active?

3. Can you briefly describe the area in which this organization is active?

4. What is your position at this organization and what does this position entail?

5. Is Artificial Intelligence applied within your organization?

6. Is Artificial Intelligence applied within your marketing department?

7. How are you involved in the use of Artificial Intelligence (within marketing) within your organization?

8. How does your organization implement Artificial Intelligence within their marketing activities?

9. How does your organization implement Artificial Intelligence within their online advertising campaigns?

10. Why did your organization choose to apply Artificial Intelligence within its marketing activities/online advertising campaigns?

11. In what ways do other companies (competitors) use Artificial Intelligence within their marketing activities/online advertising campaigns?

12. What do you consider to be the strengths and opportunities of using Artificial Intelligence in online advertising campaigns?

13. What do you consider to be the weaknesses and threats of using Artificial Intelligence in online advertising campaigns?

14. How would you describe unintended outcomes in Artificial Intelligence? From a marketing perspective.

15. Have unintended outcomes ever occurred within online advertising campaigns?
    a. If so, what unintended outcomes were these?
    b. If not, how was this prevented?

16. What measures has your organization taken to prevent unintended outcomes?

17. If that unintended outcome is detected, what steps/procedures does your organization take to repair any damage?

Interview Structure, Question 2 (Technology)

2. **To what extent does the use of Artificial Intelligence lead to unintended outcomes in online advertising?**

1. What is your name?

2. What is the name of the organization for which you are active?

3. Can you briefly describe the area in which this organization is active?

4. What is your position at this organization and what does this position entail?

5. Is Artificial Intelligence applied within your organization?

6. Is Artificial Intelligence applied within your marketing department?

7. How are you involved in the use of Artificial Intelligence (within marketing) within your organization?

8. What definition do you give to Artificial Intelligence?

9. What strengths and opportunities do you see in the use of Artificial Intelligence?

10. What weaknesses and threats do you see in the use of Artificial Intelligence?

11. How would you describe unintended outcomes in Artificial Intelligence? From a technological perspective.

    a. How would you define a fair and unfair outcome?

12. What causes would you identify for unintended outcomes in Artificial Intelligence?

13. Can you describe a situation where unintended outcomes in Artificial Intelligence had a positive outcome? And can you describe a situation where unintended outcomes in AI have had a negative outcome? In particular, unintended outcomes in the field of advertisements.

14. How could unintended outcomes in Artificial Intelligence be prevented? In particular, unintended outcomes in the field of advertisements.

Interview Structure, Question 3 (Legal)

3. **How are companies that are using Artificial Intelligence for their advertising campaigns currently meeting (legal) framework(s) in the EU?**

1. What is your name?

2. What is the name of the organization for which you are active?

3. Can you briefly describe the area in which this organization is active?

4. What is your position at this organization and what does this position entail?

5. Is Artificial Intelligence applied within your organization?

6. Is Artificial Intelligence applied within your marketing department?

7. How are you involved in the use of Artificial Intelligence (within marketing) within your organization?

8. Are you familiar with the FACT regulation?

9. To your knowledge, what measures are currently regulated by your company /the EU to ensure that the AI campaigns function according to the FACT regulation? FACT: (Fair, Accountable, Confidential, Transparent)

10. Can you share with me how these measures work/do not work? And why?

11. How would you describe unintended outcomes in Artificial Intelligence? From a legal perspective.

12. Can you share with me which (other) measures are currently being taken by your company /the EU to prevent unintended outcomes?

13. Do you agree/disagree with stricter controls on the use of Artificial Intelligence in online advertising campaigns? And why?

14. What measures do you think should be adopted or omitted in the future to support/promote fair AI? And why?

Interview Structuur, Question 2 (Marketing)

1. **Hoe implementeren organiseren Artificial Intelligence nu binnen hun online advertentiecampagnes?**

   1. Wat is uw naam?

   2. Wat is de naam van de organisatie waarvoor u actief bent?

   3. Kunt u kort beschrijven op welk gebied deze organisatie actief is?

   4. Wat is uw functie bij deze organisatie en wat houdt deze functie in?

   5. Wordt er Artificial Intelligence gebruikt binnen uw organisatie?

   6. Wordt er Artificial Intelligence gebruikt binnen uw marketingafdeling?

   7. Op welke manier bent u betrokken bij het gebruik van Artificial Intelligence (binnen marketing) binnen uw organisatie?

   8. Op welke manier implementeert uw organisatie Artificial Intelligence binnen hun marketingactiviteiten?

   9. Op welke manier implementeert uw organisatie Artificial Intelligence binnen hun online advertentiecampagnes?

   10. Waarom heeft uw organisatie ervoor gekozen om Artificial Intelligence toe te passen binnen de marketing activiteiten/online advertentiecampagnes?

   11. Op welke manieren andere bedrijven (concurrenten) Artificial Intelligence binnen hun marketing activiteiten/online advertentiecampagnes?

   12. Wat beschouwt u als de sterktes en kansen van het gebruik van Artificial Intelligence binnen online advertentiecampagnes?

   13. Wat beschouwt u als de zwaktes en bedreigingen van het gebruik van Artificial Intelligence binnen online advertentiecampagnes?

   14. Hoe zou u 'onbedoelde uitkomsten' in Artificial Intelligence beschrijven? Vanuit het marketing perspectief.

   15. Hebben zich wel eens onbedoelde uitkomsten voorgedaan binnen online advertentiecampagnes?

       a. Zo ja, wat voor een onbedoelde uitkomsten waren dit?

       b. Zo niet, hoe is dit voorkomen?

   16. Welke maatregelen heeft uw organisatie genomen om onbedoelde uitkomsten te voorkomen?

   17. In het geval dat er onbedoelde uitkomsten gedetecteerd zijn, welke stappen/procedures neemt uw organisatie dan om eventuele schade te herstellen?

Interview Structuur, Question 2 (Technologie)

2. **Op welke manier leidt het gebruik van Artificial Intelligence in online advertentiecampagnes tot onbedoelde uitkomsten?**

1. Wat is uw naam?

2. Wat is de naam van de organisatie waarvoor u actief bent?

3. Kunt u kort beschrijven op welk gebied deze organisatie actief is?

4. Wat is uw functie bij deze organisatie en wat houdt deze functie in?

5. Wordt er Artificial Intelligence gebruikt binnen uw organisatie?

6. Wordt er Artificial Intelligence gebruikt binnen uw marketingafdeling?

7. Op welke manier bent u betrokken bij het gebruik van Artificial Intelligence (binnen marketing) binnen uw organisatie?

8. Welke definitie geeft u aan Artificial Intelligence?

9. Welke sterktes en kansen ziet u in het gebruik van Artificial Intelligence?

10. Welke zwaktes en bedreigingen ziet u in het gebruik van Artificial Intelligence?

11. Hoe zou u onbedoelde uitkomsten in Artificial Intelligence beschrijven? Vanuit het technologie-perspectief.
    a. Hoe definieert u een eerlijk en oneerlijke uitkomst?

12. Welke oorzaken zou u identificeren voor onbedoelde uitkomsten in Artifical Intelligence?

13. Kunt u een situatie beschrijven waarin onbedoelde uitkomsten in Artificial Intelligence een positief gevolg hebben gehad? En kunt u een situatie beschrijven waarin onbedoelde uitkomsten in AI een negatief gevolg hebben gehad? In het bijzonder, onbedoelde uitkomsten op het gebied van advertenties.

14. Hoe zouden onbedoelde uitkomsten in Artificial Intelligence voorkomen kunnen worden? In het bijzonder, onbedoelde uitkomsten op het gebied van advertenties.

Interview Structuur, Question 3 (Wetgeving)

3. **Hoe gaan organisaties die gebruik maken van Artificial Intelligence binnen hun online advertentiecampagnes om met de Europese wetgeving?**

1. Wat is uw naam?

2. Wat is de naam van de organisatie waarvoor u actief bent?

3. Kunt u kort beschrijven op welk gebied deze organisatie actief is?

4. Wat is uw functie bij deze organisatie en wat houdt deze functie in?

5. Wordt er AI gebruikt binnen uw organisatie?

6. Wordt er AI gebruikt binnen uw marketingafdeling?

7. Op welke manier bent u betrokken bij het gebruik van Artificial intelligence (binnen marketing) binnen uw organisatie?

8. Bent u bekend met de FACT – regulering?

9. Naar uw weten worden er maatregelen gehanteerd er nu door uw organisatie/de EU gehanteerd worden om ervoor te zorgen dat de AI-campagnes volgens de FACT-regulering functioneren? FACT: (Fair, Accountable, Confidential, Transparent)

10. Kunt u met mij delen op welke manier deze maatregelen wel/niet functioneren? En waarom?

11. Hoe zou u onbedoelde uitkomsten in Artificial Intelligence beschrijven? Vanuit het wetgeving-perspectief.

12. Kunt u met mij delen welke (andere) maatregelen er nu door uw organisatie/de EU gehanteerd worden om onbedoelde uitkomsten te voorkomen?

13. Bent u het wel/niet eens met strenger controleren van het gebruik van Artificial Intelligence binnen online advertentiecampagnes? En waarom?

14. Welke maatregelen vindt u dat er in de toekomst aangenomen of gelaten moeten worden om eerlijke AI te ondersteunen/te stimuleren? En waarom?

APPENDIX C: GIOIA METHOD SCHEDULE

Gioia Method - Analysing Interview1st Order Concepts	2nd Order Concepts	Aggregate Dimensions
I would say that AI is software that does things which we previously thought could only be done by reasonably intelligent people	Definition	Artificial Intelligence
Carrying out processes by a machine replacing human or mental activities.		
In terms of systems that can reason, there isn't some sort of meaning that emerges from an interaction or reason that can emerge from an interaction between a system and a human.		
AI kan worden gebruikt voor elk systeem dat kan leren en zijn prestaties in de loop van de tijd kan verbeteren zonder dat het daarvoor expliciet is geprogrammeerd.		
AI zelf is de is de technologie. Maar het systeem dat het in werking gaat brengen is een sociaal technisch systeem. Je kijkt zowel naar de technische als sociale subsystemen en de interactie daartussen.		
Ik beschouw het als inderdaad computergestuurde systeem, gestuurde intelligentie die natuurlijk ooit in het beginsel wel een bepaalde input nodig hebben gehad van de menselijke kant, maar op basis van data en gegevens die die systemen dan verzamelen en ook een eigen logica daarop ontwikkelen.		
in feite gewoon een mens precies nauwkeurige snelle Statistiek en dat dat zo precies is, dat is heel knap, dat doet daar geen afbreuk aan te doen		
Een vorm van intelligentie waarmee machines, software en apparaten zelfstandig problemen oplossen. AI- systemen volgen algoritmen, herkennen patronen en kunnen zelf beslissingen nemen of voorstellen.		
Met behulp van AI kunnen processen geautomatiseerd worden en mensen of hun data in categorieën worden onderscheiden.		
AI door de Wetenschappelijke Raad voor het Regeringsbeleid onlangs een systeemtechnologie is genoemd. Daarmee wordt bedoeld dat het, net als bijvoorbeeld elektriciteit, een grote impact heeft op de samenleving, die vooraf niet kan worden voorzien. Om een systeemtechnologie als AI in te		

bedden is een integrale aanpak vereist, met intensieve betrokkenheid van de overheid.

So even if you have very smart people working on something, they get tired. So, whenever something is repetitive and I think there's a big opportunity for AI, I could help us eliminate or mitigate biases to improve transparency. The opportunities in Artificial Intelligence lie in its supportive elements. It allows teachers to save time in which they can invest in their students. In contrary to humans, it cannot get tired. Nou, ik denk sterk, at is zo gewoon het verwerken van enorme aantallen informatie, waardoor betere beslissingen kunnen worden genomen dan als een persoon alleen maar met de hand dingen zou doen. snelle, efficiënte werken Automatisering van repetitieve taken. Convenience is a very big one. So, you have voice assistance, for instance, and you know, you can say something to them. Even your iPhone, like your face ID, those types of things. Everything's about convenience. Sinds we AI hebben opgepakt hebben we op vele vlakken (voornamelijk advertising en seo 'copywriting') een enorme efficiëntie & kwaliteit slag kunnen realiseren. Hierdoor zie ik nog vele andere kansen om AI te betrekken in meerdere processen binnen het bedrijf. Binnen advertising b.v. zit de efficiëntie erin dat de AI dag en nacht 'doorwerkt'. Als dezelfde taken moeten worden uitgevoerd door een persoon dan hebben we minimaal 5 fulltimers extra nodig om dezelfde kwaliteit te kunnen handhaven.	Efficiency	Opportunities
I think it can help due processes like. Deal with weak amounts of data gathered, weak amounts of data deal with big amounts of data, process them, organise them, and give you possible interpretations and find possible linkages between patterns that humans would not be able to see. Machines can find that we would completely miss the point on. I would say in many disciplines that deal with big amounts of data, there's a lot of room for improvement, especially in finding patterns that have been missed	Optimization	

throughout history, geography, transportation, and organization of organizational processes. With Artificial Intelligence, the sky is the limit.

The weakness of the system that we should take into account, is that there should still be a human in the loop and that we should not just blindly trust this AI-based system.

The errors in decision-making.

I think that's a little bit more perhaps in a commercial context where, you know. Usually, companies are driven by profit motivations and especially in a more company commercial context. That's not to say that you're not going to solve preference for your own or your partner's content versus some small business that is not very profitable for you. So, these things are very difficult to prove or detect in the first place.

Something was because these systems are so complex and at some point, the systems' designers don't know what's going to come out at the other end. So that's a very big challenge for a legislator as well to address that. And the companies themselves, because they say, well, how can we provide explainable AI if we don't know ourselves?

Great power leads to great damage, for example, The Childcare Benefit Fraud Scandal.

Weaknesses

AI can consist of a black box. If developers do not know why certain decisions are being made this can be a threat.

So, I think the biggest interest of intelligence is that we don't understand it. The people that use it often to understand it, the people who receive it, don't understand or don't even know that they're being subjected to it.

AI (vooral deep learning) is nog grotendeels een black box. Het is dus vrijwel onmogelijk om te bepalen waarom er bepaalde keuzes gemaakt worden door de AI. Dit kan leiden tot onbedoelde uitkomsten.

AI kent geen grijs gebied het heel zwart-wit ja zeker, en het grijze gebied in de menselijkheid is datgene wat ons verbindt en wat ons ook de mogelijkheid geeft om fouten te kunnen maken.

Black Box

So, in practice, most uses of artificial intelligence today are negative because they don't respect human rights either in the process or in the result. So, either the way they are deployed or gather data, it's done for control purposes, surveillance purposes or profit.

Human Rights

Mensen worden door AI in een hokje geduwd waardoor ze allerlei bepaalde dingen wel zien op het internet en andere dingen niet, waardoor ze echt heel erg een bepaalde manier van denken kunnen krijgen zonder dat ze andere denkbeelden ook afwegen omdat ze die gewoon letterlijk niet te zien krijgen. (Polarisatie)
Ik denk dat de bedreigende in vooral in zit dat je op een gegeven moment niet meer ziet wat erdoor AI beïnvloed is en wat niet.

Threat

Je kan de doorsnee burger niet kwalijk nemen dat hij dat niet snapt of niet ziet of überhaupt begrijpt. Maar daar zit wel een probleem en ik denk dat je d'r ja, daar moet wel iets gebeuren en bij wie dat bij die verantwoordelijkheid ligt, is onduidelijk
Digitalisering en mensenrecht geworden, met het idee dat AI, maar ook digitalisering in brede zin de rechten van de burger op bepaalde manieren schendt als het gaat om discriminatie. Bijvoorbeeld, als het gaat om rechtsbescherming, hoe weet je wat voor een besluit er over jou is genomen, als je slecht bij je eigen informatie kan?
Schaalbaarheid, één kleine keuze van AI kan grote gevolgen impact hebben.
Continuïteit, een AI-systeem maakt de keuzes waar het voor geprogrammeerd is.
Intransparantie: En nou ja, wat het wat ons betreft betekent is dat het nog steeds geen probleem is, want ik zie het eerder als een onderdeel van het probleem, namelijk de intransparantie van dit soort systemen zorgt ervoor dat jij ook nooit een klacht zal indienen.
Je krijgt niet volledige informatie krijgt dus dat door of ritmes ervoor wordt gezorgd dat jij maar bepaalde werkte informatie krijgt, waardoor je daar je manier van denken op gaat baseren in plaats van een meer compleet beeld.

Er zitten nadelen aan AI, maar deze kunnen makkelijk getackeld worden. Een voorbeeld is dat AI-tijdsbesparing oplevert. Door tijdsbesparing kunnen er minder uren gerekend worden wat weer tot minder inkomsten zou kunnen leiden waardoor het businessmodel onder druk komt te staan.

Door minder tijd te besteden aan het praktische taken, zoals het verhogen of het verlagen van een klikprijs, besparen we heel veel tijd. Deze tijd kunnen wij nu besteden aan de strategie, creativiteit en contact met de klant waardoor de klant van een beter resultaat en ervaring geniet. Dit biedt daarom geen bedreiging, maar kansen.

Door AI hebben we juist een opschaling kunnen maken in het aantal uren voor klanten en zijn wij juist gegroeid.

it makes it harder for the ones that are subject to it to find redress, to know what's happening, to ask for avenues for redress, and for judges to also be able to trace back in and give a good result in court proceedings.

That's part of the problem because people do not know that they are being discriminated against while they are being discriminated against. And when we are revealing the system that's behind us, we then notice, oh, there is discrimination, but people just didn't do it a machine did.

So, if you think about advertising campaigns more specifically, usually you think about a few main players. So, you have Google and Google ads and then you have Facebook. And there is a saying that no single ad goes through the Internet and reaches the consumer or user without Google or Facebook being involved at some point.   Marketing

One of the main ways that display ads is being delivered. So, you're not in a closed ecosystem. You're in a far more open, open ecosystem.

 And it's an automated auction. So, we use data, we try to reach very automatically the users and connect them with the content that you want to show. So, it's very granular. So, it's moving increasingly towards this consumer-centric advertising. It's very much this automation, but also AI-based because to identify the right user group or user, you have to analyse all this data and make a connection there. So, this is pretty much why it's very

hard for me to see any marketing campaign of any scale or even a small one. Because if I'm a small company or an individual, I imagine I'll go to one of these companies and just set up a campaign. And usually, different tools help me to identify how I reach my audience, or at least, you know, have this idea that I'm going to reach the audience. So, the marketing companies are talking about I kind of have the feeling that there are intermediaries in between. But the end goal is to go to the end user. Not to the ecosystem pretty much.

Doordat Google vier jaar geleden overstapte op het gebruik van AI ging Team Nijhuis mee met deze shift om hun marktpositie te kunnen behouden. Google stapte vier jaar geleden af van hun dagelijkse CPC-model. Een model waar Team Nijhuis zich mee bezighield en wat voor hen een Unique Selling Point was.

Doordat Google overstapte op het gebruik van Artificial Intelligence, stapte Team Nijhuis ook over op deze praktijken om hun marktpositie te behouden. Vooroplopen en meegaan met nieuwe ontwikkelingen. Dit brengt fouten met zich mee. Maar dit houdt ook in dat je bij nieuwe technologische ontwikkelingen deze direct gaat doorvoeren zodat je deze kunt aanbieden aan de klant.

Deze technologische ontwikkeling (test case) worden per klant getest. En vanuit door wordt besloten of er meerdere klanten aangehaakt moeten worden en het idee voor alle klanten moet gaan werken.

Op deze manier kunnen we successen voor klanten inzetten. Het uiteindelijke doel is om de klant de beste klantervaring bieden.

Tig procent dat jij daadwerkelijk mannen bereikt, als je nog even uitgaat van twee genders maar dan, ja, dan ben je dus eigenlijk gewoon aan gokken je begint in van je bereikt net zoveel vrouwen of mannen leeftijd. Is de kans 25 procent dat je nog dat werkelijk wel leeftijd bereikt, die je aan de targeten bent, omdat ze het zelf ook niet meer weten, omdat mensen steeds inderdaad dat delen. De wetgeving wordt steeds strenger. Dus wat we daarin wil, ook in weg proberen op marketinggebied is om de data zoveel mogelijk naar je eigen infrastructuur op te halen en naar je eigen bronnen of je eigen

Why employ AI within marketing?

systemen. Dus stel dat je geen dit expert in je was boek en je hebt in stuk en je hebt allemaal een beleid campagnes draaien. Laat dan die data niet daar staan, maar halen naar jezelf op, zodat je zelf eigenlijk gaat leren over je klanten, dus dat je kan zien. Dat zou het mooiste zijn als je t ook kan koppelen met je bestelsysteem bijvoorbeeld dat je kan zien. Oké, deze persoon heeft een fiets besteld. Welke content heeft ie allemaal bekeken met heeft, en dat je dat zelf kan doen zonder dat je daar dus derde-partij software voor gebruikt.

Wij doen de Artificial Intelligence analyses voor de klant. Google Analytics stopt per 1 juli 2023, dus dit houdt in dat organisaties die jarenlang aan data daarin opgeslagen hebben deze data kwijt zijn. Dit geldt met name voor MKB-bedrijven. Door het verlies aan data verlies je ook capaciteit om Artificial Intelligence toe te gaan passen.

De boodschap die wij meegeven is: Koppel externe bronnen, zoals CRM-systemen, webshops, interne bronnen, kadaster gegevens etc. Hier kun je vervolgens analyses op doen en de inzichten hiervan teruggeven aan de onlinekanalen. Hierdoor, kun je veel gerichter targeten.

Wij doen leadgeneratie radar, marketing, zoals radar, marketing en het ook weer wel marketing. En daar zitten natuurlijk allemaal vormen van achter op dat soort plot. Platformen om content, die je dus als het ware beschikbaar stelt naar doelgroepen ja te presenteren, en dan uiteraard met bepaalde budgetten en dergelijke, zorgen dat je de juiste ja clickrate en conversie krijgt Google werkt met biedstrategieën die gestuurd worden door Machine Learning

Toen Google overstapte naar AI ging Team Nijhuis ook aan de slag met AI en stelt zichzelf de volgende vragen: Hoe gaan wij deze biedstrategieën wanneer wel of niet gaan inzetten? En hoe werken deze?

Tweede mind shift: We moeten deze biedstrategieën voor alle klanten inzetten. Want Machine Learning traint zichzelf. Hoe meer input (data) Machine Learning krijgt hoe beter de resultaten uiteindelijk zullen worden.

For what tasks will you employ AI within marketing?

Nu is Team Nijhuis volledig afhankelijk van de door Machine Learning gestuurde biedstrategieën.
We use Artificial Intelligence mainly for executive marketing. Meaning, that we focus on executive marketing activities. This can be performance: How can improve the current marketing activities? Or find new customer segments. This is done through clustering.
Artificial Intelligence can be used for Search Engine Optimization, in which we refocus on the search terms and try to find out which search terms have been the most successful. In this case, we focus on a backwards customer journey.
For example, A customer has just ordered a product on our website. Which advertisements and search terms did he or she interact with?

Nou, dat is wel het gros, ook het meest belangrijke en uiteraard de advertenties die wij zelf maken en plaatsen. Hé, dus die inkoopkant die moeten ook echt aan allerlei Eisen voldoen. En wat wij doen is ook allerlei testen uitvoeren. Dus voordat wij een nieuwe campagne opzetten hé, testen wij het Oké, hé, hoe ziet het er nu uit? Voldoen we aan de policies over intellect? Hé dat niet in strijd mag zijn met intellectueel eigendomsrechten niet in strijd mag zijn met oneerlijke reclame. Nou, noem maar op	How do companies prevent unintended outcomes	
Unintended outcomes are where things happen that were not intended by ever designed, planned, or implemented in AI. These unintended outcomes are if the tool just doesn't do what it should, what it's supposed to do that it should be better prepared to learn from real cases. Het komt regelmatig voor dat men een bepaalde input in een algoritme zendt en er een ander gepland output uitkomt. Depends on the context in which the AI system is used. you have a predetermined goal, but the way you reach the goal is not entirely determined. So perhaps in the way of reaching that goal, you end up having these externalities	Definition unintended outcomes	Technological Unintended outcomes

Bij het genereren van content kunnen er 'gekke' of onlogische teksten ontstaan. Ook al is de AI goed in staat menselijke taal te interpreteren en te genereren. Toch heeft het een vrij kort geheugen om langere connecties te maken. D.w.z., tekst die eerder is genoemd kan al snel worden 'vergeten' of dat er verkeerde conclusies uit getrokken worden. Om deze rede moet er altijd nog een mens aan te pas komen voor de eindcontrole. Het advertising stuk ligt wat anders, hier voert de AI wel aanpassingen door zonder menselijke controle. Dit kan omdat het spectrum waarin de AI mag handelen heel klein is, en voldoende getraind om aanpassingen correct door te voeren.

	Data	Causes unintended outcomes

I would say that in 90% of the cases it is because of the data because we use ever larger data sets to train ever larger models. This brings us to a point where it is almost impossible to check the data to ensure that there are no harmful sentences or images or whatever in there.
I think it's this often also it's like biased databases that are Yeah, it's I think mostly it's lack of care like I think it's rarely malintent but it's incomplete databases very often made up of very small demographics or very specific demographics and needs that nobody's looking for these things exposed either.

And then I think it's also how the results are used. So, you know, there are things like. How do you handle uncertainty? So, when the algorithm says, okay, I'm 60% sure that it is a but also there's a 40% per cent chance of B, How do you handle that? Do you say, okay, then I take A? Or do you say, okay, I'm not sure
If information is leaked, if you have a model that was trained on your data and if later on by inspecting that model, we can learn that specific individuals belong to specific political groups, then that can be a source of discrimination down the line.
And I would say if we don't have enough training data if we don't have enough information on the process itself, and this is, for instance, now we

could also the tool is ready, we could put it to the University of 20, but it doesn't make sense because we did not test it enough with training data
Wanneer de dataset waarop de ai getraind is niet compleet is, of niet op orde dan is het nadat de AI is deployed lastig/vrijwel onmogelijk te achterhalen waar het fout gaat.
Niet de tijd nemen om t data te exploreren zoals bij de Toeslagenaffaire gebeurd is.

Je hebt al heel snel eigen vooroordelen en aannames. Deze verwerk je als marketeer weer in de systemen.   Prejudices/Vooroordelen
It's not a secret that humans are biased, that many institutional agencies are virtually biased and most often are biased in the same direction. They're either racist, they are sexist, they are classist, they're very, very disciplinary. They're blind sometimes to the specificities of people. So, it's hard to claim this would not have happened with Artificial Intelligence.

Neurale netwerken hebben ook nog de reputatie van een black box, waardoor   Black Box
er dingen kunnen gebeuren die niet gewenst zijn
We beginnen juist ook zo simpel om mensen een beetje aan de hand mee te nemen en ook heel goed te kunnen leren wat er goed gaat, wat er fout gaat als je meteen begint met een black-box en die lening al greep. Maar ja, dan weet je zelf ook niet hoe die tot z'n beslissingen komt en dat is natuurlijk in de algemene zin.

Dus een simpele beslisboom of een lineaire regressie die heel erg uit te   Modelling
leggen is ook om de klant eigenlijk mee te nemen. Want veel van onze klanten zijn nog niet heel volwassen op datavlak die zijn echt de eerste stapjes aan het zetten. Dan begint het ook vaak nog met uw hou mans terugkijken naar je performen omdat ze die inzicht ook nog niet altijd hebben en dan de diagnosticeren zeg maar, kunnen stellen over waarom stijgt de omzet of om dat soort vragen. En dan daarna eigenlijk een stap gaan zetten naar oké, kunnen we misschien hele kleine user cases identificeren die

impact hebben op een bedrijfsdoel zoals behoud van klanten of zo. Misschien kunnen we wel klanten gaan identificeren op basis van hele simpele regels, die misschien wel weggaan of waar het relevant is om ze nog eens een keer te bellen, of Normaal dus eigenlijk om ze ook klantmanagers te gaan voeden met informatie, om dan toch wel aan te tonen dat dat impact heeft, terwijl je dan, ja eigenlijk maar minimale data gebruikt en ook een minimaal voorspelmodel aanmaken met eigenlijk.

	Data	Prevention unintended outcomes
The first step is given that I believe most of them come from the data to just check your data better to make sure your data comes from what it contains. And then I think also the explainability part, the transparency part. So actually, understanding what your algorithm does can help. When in a testing phase of an AI system the outcomes are discriminative then you could change either the cohort of users or the testers, for instance, so that you diversify the data. Vooroordelen kun je ook weer makkelijk ontkrachten door het gebruik van data. Heel moeilijk en ja, big Data speelt natuurlijk een enorme rol, dus wil je ook iets met een bepaalde zekerheid kunnen gaan, ja, hé, voorspellen of beredeneren en in dit geval herkennen, dan heb je gewoon heel veel data nodig. De wetgeving wordt steeds strenger. Dus wat we daarin wil, ook in weg proberen op marketinggebied is om de data zoveel mogelijk naar je eigen infrastructuur op te halen en naar je eigen bronnen of je eigen systemen. Dus stel dat je geen dit expert in je was boek en je hebt in stuk en je hebt allemaal een beleid campagnes draaien. Laat dan die data niet daar staan, maar halen naar jezelf op, zodat je zelf eigenlijk gaat leren over je klanten, dus dat je kan zien. Dat zou het mooiste zijn als je t ook kan koppelen met je bestelsysteem bijvoorbeeld dat je kan zien. Oké, deze persoon heeft een fiets besteld. Welke content heeft ie allemaal bekeken met heeft, en dat je dat zelf kan doen zonder dat je daar dus derde-partij software voor gebruikt.		

Dus als dat gewoon niet geschikt is om AI te gebruiken, of je hebt überhaupt geen probleem of uitdaging die je wil oplossen, dan zou ik het gewoon niet gebruiken. Je moet wel een doel d'r mee hebben.	Business understanding
I think focusing on transparency is also an opportunity to observe what's going on. So, kind of increasing the transparency on the market. So not only me as a consumer, but I'm also seeing that this is why I get this ad.	Transparency
And another thing I think is very important that you just mentioned is to have a continuous evaluation. Comparing the AI system with the outcomes that a human (teacher) would create and compare of these outcomes contain huge differences. And a lot of testing before the tool is used. Dit proberen we te voorkomen door zoveel mogelijk van dit soort algoritmes te testen/experimenteren. Als je niet experimenteert kun je ook niet uitsluiten. Als een uitkomst niet uitpakt zoals jij verwacht, is dit nog steeds een goed resultaat want je kunt het meenemen in je aanpassingen. Waardoor je weer meer focus aanbrengt in je huidige strategie. Want je weet dit stukje werkt niet voor deze klanten en voor deze setting. Hierdoor kun je makkelijker je algoritme hervormen. Google adviseert daarom om standaard 10% van je advertentiebudget te gebruiken voor experimenten met Artificial Intelligence in onlineadvertenties. Look at the context where such an AI can be implemented, where it could be maybe less discriminatory.	Evaluation
Making sure that all the opinions are taken into account. This is something important to have all stakeholders involved. Generate dialogue with a broad set of stakeholders so that the impacts of AI can be brought to the table from very diverse perspectives.	Representation stakeholders
Always still keep the human in the loop.	Humans in the loop

Wij houden reviews over producten en systemen en innovaties. Nieuwe diensten gaan door een gehele review van door mensen die dus wel gespecialiseerd zijn en nadenken over wat on intens consequenties zijn, die gespecificeerd zijn en meer data zetten en voorzien. Is dat wel breed genoeg? Heb je rekening gehouden met minderheden wil je niet juist die mensen iets meer representeren zodat je gelijke en uitkomst krijgt?
, the weakness of the system that we should take into account, that there should still be a human in the loop, that we should not just blindly trust this AI-based system
By imposing human oversight, especially in high-risk applications, I think the need to have humans taking part in the algorithmic decision loop can prevent discrimination from occurring and directly imposing that outcome should not be non-discriminatory. – Fernando
Wij geven de ai een heel smal kader waarin deze mag handelen. En controleren de resultaten om de dag. Op deze manier kunnen wij voorkomen dat er dingen uit de hand lopen

There might be options for changing policies around or changing frameworks within which infrastructure the AI should or should not be included or implemented. EU high-level expert groups framework on trustworthy AI which includes seven ethical principles on how to create systems that would be desirable in society.	Regulation
Is it enough if the outcome is fair or just Also the process needs to be fair. So, an AI that uses an unfair process can still come to a fair result by accident or by design. So, I would say there's even that differentiation to make it. And other than that, I would say a fair outcome is something that is within, of course, the laws as kind of the minimum standard, but also what is socially acceptable ik denk dat dat dat je dat eerlijk en oneerlijk dan heel erg ziet vanuit het oogpunt van de persoon	Fair and unfair outcomes

als er technische fouten in zitten, waardoor de gewoon een beslissing komt die uitkomt die je niet zou verwachten, kom jij op een or oneerlijke beslissing uit. Het kan ook model biased is, waardoor bepaalde groepen mensen meer benadeeld worden dan andere mensen. Dat lijkt me ook oneerlijk.

Regulation positively impacts AI by creating incentives to advance tools to create AI that is not biased and prevents discrimination.	Opinion future regulations	Legal

I always find it good if there are more rules. I would say more rules to make sure that nothing happens, especially if it's real data like student grades.
Ja, tuurlijk ben ik voorstander van, want ik vind het echt heel belangrijk dat er nou, ja, dat gaat om eerlijke advertenties, dat niet verwarrend mag zijn,

Want net als bij de AVG kan het soms ook doorslaan vind ik tuurlijk. Het is heel belangrijk dat alle bedrijven zich bewust zijn van oké, hoe ga je om met persoonsgegevens en aan de andere kant denk ik ook van: ja, we moeten niet doorslaan dat het juist weer gaat beperken en dat is dat geldt eigenlijk voor die AI ook heel goed dat het wel gereguleerd wordt, maar pas op, rem, niet de innovatie daarin. En nou ja, dat zijn wel twee aspecten tussen ook één bescherming en innovatie dat zich dat niet, nou ja, in elkaar zwaar water gaan lopen.
I notice also that there are many aspects where I would rather be stricter beforehand than afterwards when something happened. So better to have a bit stricter rules beforehand and nothing happens than to be a bit more lenient and say: "Yeah, we'll be fine."
So technological innovation needs to find a place and needs to be fostered because it's from the perspective of economic growth important.
I think it's positive that they and there is room for developing such regulations, but then those regulations would account more for the network ness of I that is not just the system but like we both people and systems.
So, there is a positive side to the AI ACT. But on the other hand, if you only focus on the system, if you only focus on the AI itself, you might lose sight of the broader societal and ethical implications

Well, is there one specific regulation that can address that or is there one? And Yeah, no, that's not exactly how this works, right? So, if you want to protect or not protect, but if you want to increase fairness, and contestability in markets, if you want to have more competition, you're going to look at that specific area.

So, I think we're going to have to. What I like about this kind of principled approach is that you're trying to at least set the value straight and then adjust the kind of because with values your kind of setting what do we want our society to look like?

So, we wanted flair. And then if it does not match our values, we can do something about it. And this is the first step. And these different legal fields, perhaps one thing is that you want them to share those values. That's one of those aspects where now there's a lot of conflicts and you know.

Het is goed te kijken naar het risico van systemen en op basis van een risiconiveau bepalen welke aan welke regelgeving het dan moet voldoen, welke voorwaarde is, en dat er ook sommige systemen gewoon niet gebruikt moeten worden.

En we hebben gewoon te maken met wettelijke bewaartermijn en die maken al Di. Dit hele pakket maakt onderdeel uit van die ISO-normering die daar dan weer ja toezicht op houdt of dat inderdaad controleert of we dus ook op die manier dat wij zeggen ook werken. Dus ja, bewaar termijnen voor personeel, data bewaar termijnen voor klantdata bewaar termijnen voor nou ja, eigenlijk financiële daten alles wat daarbij komt kijken en het geeft ons: ja, allereerst een voorsprong in de markt, omdat je veel makkelijker alsof je leeft, je kunt aantonen dat je he dat het niet alleen maar gesproken en geschreven woord is, maar dat het ook inmiddels een normering wordt ge, ja, geaccrediteerd en voor onszelf geeft het inderdaad ook gewoon wel een ja, een rust, rustig gevoel eigenlijk dat je weet dat je dat de basis goed is. En als er dan eventueel iets niet klopt, dat je zeker niet op nul begint, maar dat je wel heel makkelijk dat nee, hier konden we trouwens ook even niets aan doen, maar slipt er dan toch iets tussendoor, glipt er iets tussendoor, dan is het ook daadwerkelijk een uitzondering op het geheel. Het is niet dat in één

keer de hele organisatie een probleem heeft, maar oké, we hebben hier een situatie. Die kunnen we isoleren en daar actie op ondernemen.

Ik vind één ding alleen heel lastig en daar komen we misschien ook bij die wetgeving en bij de bij defect ik ben er dus in grote lijnen uiteraard mee eens en dat ga denk ik altijd een beetje voor eigen bestwil naar het aspect van: ja, het moet niet mensen onnodig buitensluiten discrimineren. We moeten ervoor zorgen dat ai misschien niet op een niveau komt dat dat beslissingen letterlijk neemt, als wat ik net zei, heel erg zwart-wit dat het uiteindelijk bewijs spreken, de oorlogsvoering gaat bepalen tussen landen of organisaties, dat het de controle krijgt over wapens en dat soort dingen. Anderzijds vind ik het ook wel heel lastig in een wereld waarin wij allemaal iets, ook wel zakelijk gezien, aan elkaar te verkopen, dat de wereld op privacy, niveau zo dicht wordt getimmerd ik vind dat een nadeel, ook weer van dit alles en niet omdat ik nou iedereen maar hé, onbeperkt wil kunnen spammen en lastig wil kunnen vallen. Maar ik merk wel dat dat vandaag de dag van ja, enorm lastig is en wordt om eigenlijk met mensen in contact te komen en dat het ook tegenwoordig wel een heel snel excuus is, inderdaad van mensen om te roepen: van hoe kom je aan m'n gegevens, hoe durf je het eigenlijk m contact met mij op te nemen?
And we can see this in the case of GDPR because it's very hard for many SMEs to comply with GDPR, whereas we act as a society. Like GDPR is something that's almost like old news like it's so basic, so average. For many SMEs, it was very hard back then, and it's still hard today to comply.
If SMEs are not involved in the process and if they're not involved in the process, they will not the registration will not fit them and it's not going to be realistic for them to implement.
the fear is that if our strength relies on SMEs not being involved, our strength in European strength if our strength relies on SMS and the SMEs are not involved in the process, who is this legislation g help? One of the main points is that SMEs are never in standard-setting bodies, and they're never asked about.

I think some kind of organization, program or committee or something offered would help.

Potential Regulations

AI, Big Data en ICT zijn gezamenlijk de drie eenheden en staan met elkaar in verband. Je kunt niet het één gaan reguleren en verwachten dat daarmee het probleem is opgelost. Je zult voor elke sector een aparte regulering en/of wetgeving moeten bedenken/opstellen.

There is a dialogue about these systems in society. That there is room for dialogue about the implications of these systems in society and where the innovators themselves are at the table because so far, it's a power imbalance. You see the large, big tech companies creating the platforms that define the way how discrimination can happen. So, they, from an economic perspective, set the rules.

So, if you look at the companies, especially big tech companies, you have very so much technical expertise there. And I think too kind of. We should perhaps invest and involve these experts in stronger

I think focusing on transparency is also an opportunity to observe what's going on.

I think there's a role for researchers to, you know, explore these avenues, and see, well, do we see any patterns or connections or what is going on there? I think advertising specifically, advertising is also evolving

Companies should not use AI because they think it is a new trend. They need a very clear business understanding before implementing it within their business practices.

Companies should start by implementing Artificial Intelligence within their business practices in small steps, such as starting with linear regression and/or a simple decision tree. This will help to understand how AI works and how it can work for them and also prevents large biases and unintended outcomes.

Via de normen van de privacywetgeving de AVG dus, ja, niet zonder toestemming verkrijgen, alleen maar aanschrijven mensen de mogelijkheid bieden om uit te kunnen laten schrijven. Kijk, met klantdata hebben wij

intern allerlei systemen, waarin wij dat dus opslaan dat moet allemaal wat we zeggen Encripted zijn. Onze organisatie voldoet dus aan allerlei beveiligingsnormen zowel Fysiek als digitaal, om ervoor te zorgen dat niet zomaar op straat komt te liggen. We hebben hele duidelijke gebruikers met protocollen binnen ons bedrijf, van wie waarbij mag en wel of niet bij kan. En dat wordt dus ook, hé, want je kunt zeggen dat je het allemaal organiseert en zo doet. Maar dat moet je dus ook meerdere keren per jaar vanwege die ISO-certificering controleren en aantoonbaar maken. Dat je dus ook zo werk en daar zijn allemaal weer triggers voor en je wordt minimaal één keer per jaar extern daarop geoordeeld. Dus daarmee is het niet alleen zeggen dat je het doet, maar je kunt ook de bewijslast leveren dat je het zo doet. Nou, anderzijds hebben wij te maken met inderdaad de data die onze klanten in onze systemen opslaan wij zijn de eigenaar van ons systeem met die software, service, maar alles wat de klant in het systeem zet en beheert en gebruik, dat blijft eigendom van onze klant. Nou, ook daarin speelt beveiliging een hoger, deze hoge prioriteit, dat houdt in. Wij werken vooral voor Europese opdrachtgevers hé, dat dataopslag alleen maar gegarandeerd in Europa mag zijn, dus binnen de eer. De datebases waarin het wordt opgeslagen en de verbindingen die tussen de datebases liggen, zijn allemaal encripted.

Maar we moeten ook beschermd worden als een burger. En nou ja, dat dat zou zo'n middel kunnen zijn dat het ja op altijd heel duidelijk is dat als je een aan aanbod krijgt of je krijgt een advertentie, dat je het eigenlijk terug zou kunnen herleiden, dat je op een knop zou kunnen drukken om te zien waar dat allemaal vandaan komt. Ja.

So overall, what we ask for in most legislations is either SME exemptions or SME support or SME exceptions.

So, for example, in conformity assessment or setting up quality management. Like there's one in for high risk. I often want the setting up and own quality management department. An SME cannot afford these, and SMEs can be two or three people. This takes a lot of money. It takes a lot of time. This takes a lot of knowledge either in-house they don't have or hiring outsiders with

money that they also don't have. So, it's just not realistic to pretend that I am going to take up in this continent if SMEs are not first asked and legislation is targeted to them and not to the companies.

More money should be made available for SMEs to implement the upcoming AI Act practises within their daily practises.

But the idea is no legislation with that money and no money with no reform without money or without investment and no investment without reform.

What kind of AI do we want and how do we make it happen? And that comes with money and that comes with involving SMEs because they are depending on the country, the 99, 90% of the companies in the country. So, what are we waiting for? Waiting for Google and all of these to comply when we can do it at home?

What it will do. To my knowledge, is that it will create a classification system and that would provide. Some sort of. And green light for certain types of systems. More assessments in terms of safety, second city for certain types of systems that are not proposed or not propagated.	AI Act	Current Regulations

The AI act is and has been developed based on the Product Safety Act at some older ones, like almost 20 years old. So really, they develop that on the EU level. So that the washing machine and the microwave are good enough and not harmful to people. However, the AI systems we talk about are not as clear cut as the washing machine and are not as clear cut as the microwave because of the A.I. system. Fuel for the air systems is human data and for the washing machine is electricity most of the time and also for the microwave, right?

De AI zou moeten verstrengen en zou veel meer uit moeten gaan van rechten van burgers in combinatie met de verplichtingen van de makers. De wet bestaat nu natuurlijk voornamelijk uit verplichtingen van de makers.

In short, there's this mismatch in terms of looking at systems as compared to those that are under the Product Safety Act.	Product Regulations

Se vraag of productveiligheid de meest logische is. AI heeft als kenmerk dat het veranderlijk is en tijdens het gebruik nog verder ontwikkelt. Zou voedselveiligheidswetgeving dan niet een betere inspiratiebron zijn?

Als je kijkt naar de GDPR bijvoorbeeld, dan zie je dat dat nog zelfs in de praktijk dat langzamerhand begint iedereen wel begint het onderdeel van dagelijks handelen te worden, dat je daar rekening mee moet houden. Dan heeft iedereen en privacy dus verantwoordelijke in-huis et cetera maar aan de Handhavingskant bijvoorbeeld, om ook ervoor te zorgen dat dat daadwerkelijk dat beleid uitgevoerd wordt, is nog steeds best wel een ding. Het digitale veld is heel lang overgelaten aan de markt en de techneuten. Het is nu pas dat er door de EU-reguleringen worden getroffen. Bijvoorbeeld, vorige week is de Digital Services Act in werk getreden. Dit is een variant van een regulering waar het op verschillende aspecten fout kan gaan. Een ander voorbeeld hiervan is het invoeren van de GDPR.
De richtlijnen opgesteld door een groep experts is mooi, maar om deze richtlijnen door te vertalen in de praktijk van bedrijven of in wetgeving en daar ook op kunnen handhaven is een ander verhaal.
Dus dat ook de grotere partijen, zoals de gel en de BING en er ook echt wel kijken aan. Oké, wij willen alleen maar zakendoen met adverteerders die daar gewoon hé die eerlijk zijn, vertrouwelijk zijn, transparant zijn. Dat zijn wel echt de waarden, laat maar zeggen die ook bij dat soort partijen heel oog in het panel staan. En dat zie je ook in al die polissen terugkomen. Dat is wel meer baren van polities want je hebt content, tracking policy nou, noem maar op en al die verschillende polities zie je dat heel duidelijk naar voren komen.

GDPR

Kijk discriminatie, onze wet op tegelijker behandeling, met algemene wet voor gelijke behandeling die functioneert op basis van een klachtensysteem hé dus dat jij nou, als jij merkt dat er een, dat er iets is gezegd of waar jij een vermoeden door hebt, dat er even te veel sprake is van discriminatie, dan ga je daar een klacht over indienen en dan lost het op die manier op. Alleen wat je merkt bij is dat klacht, een klacht, methode, niet zo goed functioneert ja,

Discrimination Law

dat is omdat je eigenlijk, hé, je kan zelf eigenlijk heel slecht te herkennen waar het probleem ligt, dus die verantwoordelijkheid geven bij de klager of degene die gediscrimineerd wordt.

En dat is precies de reden waarom, ja je ook wel ziet bij dat er in een regelgeving rondom digitalisering onderwerpen steeds vaker gekeken wordt naar het transparant het goed bouwen van dit voorwaarden vooraf stellen aan een computersysteem hé, dus dat zie je me eigenlijk al.

Het gaat bij AI niet alleen om de technologie die als instrument wordt ingezet. Minstens zo belangrijk is het verdienmodel. De grondstof van AI wordt gevormd door data. Data vertegenwoordigen een grote economische waarde. Een overheid die AI wil inzetten moet er voor waken niet te betalen met de data die zij hebben vergaard voor hun publieke taken. Dit geldt zowel voor data over onze waterwerken als persoonsgegevens.   Data

So, there are investigations ongoing. For instance, Google, when it comes to search advertising specifically, would be the dominant key main player. And then when it comes to displaying, Google and Google and Facebook have a duopoly. But just to give an example, if that's something you would want to look at, Google AdSense was the decision. Believe in 2019 from the commission. That's an antitrust decision, specific abuse of dominance.   Digital Markets Act

One challenge we still have, and this is something that it's part of the research   Challenges
that I try to do, is to develop the tools to try to anticipate the impacts of AI overall. So, there's a lot in current regulation that maybe focuses on the short term. So, if we have a certain application, how is that application now with the data that we have discriminating or not or impacting society in a certain way
Van der Linden wijst op een ander nadeel van de gekozen aanpak: het wordt een enorme bureaucratische exercitie voor ontwikkelaars en gebruikers om

aan te tonen dat de AI hoog-risico systemen veilig zijn, terwijl dat niet per se hoeft te betekenen dat deze daadwerkelijk veilig.

So, I think ultimately, we need different fairness metrics that allow allows us to look into different aspects of discrimination.
This fairness aspect is something. It's good that it comes in regulation or is already out there as regulation and accountability.

Fairness

I often have this discussion with the lawyers that lawyers are personally responsible for what they do. So, if they give you bad advice, you can sue them. They are liable for it. But computer scientists put out data versions with algorithms. They are not sure what exactly they are doing, and they are not accountable for it. So, accountability is a big problem.

Accountability

So, I assume that privacy violations and tracing back certain algorithmic decisions and leaking information about specific individuals can be
So, for all AIs that produce something, it could, of course, happen that they reproduce stuff from data they were trained on and that could violate confidentiality or other rights

Confidentiality

And then also transparency even for open-source software or whatever it is, in theory, open source. But then we download it from somewhere in some compiled version, and we still cannot be sure whether this is still exactly doing what we think it is doing. So that is also, I think, a problem
It is then necessary that the algorithm is transparent. So, only if we understand that why, why it did we even first have to notice that it did something other than we expected. And for that, there needs to be some kind of transparency and explainability to the algorithm.
In terms of human rights, we need to protect certain groups based on for example their political persecution (a protected characteristic).
And there you have transparency requirements. And this is where advertising comes in as well, because, of course, there are different types of advertising.

Transparency

But what we as. At this point think, well, if we let people know what's going on if then it should be okay. Most of what we're focusing on when it comes from an individual user's perspective is increasing transparency requirements and simply giving the users at least knowledge like, why did I get this ad? What do you know, what am I watching? I'm watching a deepfake and it's saying something so I should know it's a deep fake.		
Maar dan zit je ook bij de indirecte discriminatie, problematiek van zelfs als jij d'r indirect achter komt, ja, dan is maar zeker als je de indirect doet, met de bedoeling om uiteindelijk onderscheid te maken tussen mannen en vrouwen. Ik vind er wel riskant.	Discrimination	Discrimination
I'm thinking about this famous Netflix competition example whereby looking into the data that was made available by Netflix, people could learn about sexual orientation and political preferences of individuals. And these are two elements that allow them to categorise individuals along protected characteristics. And of course, then it depends on how that information is used. But that can create a basis for discrimination	Example Discrimination	