# Handelshøyskolen BI

## GRA 19703 Master Thesis

Thesis Master of Science 100% - W

### Predefinert informasjon

| | | | |
|---|---|---|---|
| **Startdato:** | 16-01-2022 09:00 | **Termin:** | 202210 |
| **Sluttdato:** | 01-07-2022 12:00 | **Vurderingsform:** | Norsk 6-trinns skala (A-F) |
| **Eksamensform:** | T | | |
| **Flowkode:** | 202210\|\|10936\|\|IN00\|\|W\|\|T | | |
| **Intern sensor:** | (Anonymisert) | | |

### Deltaker

| | |
|---|---|
| Navn: | Lizeth Larsson Brandsås og Charlotte Vaktel |

### Informasjon fra deltaker

| | |
|---|---|
| **Tittel *:** | How Can Artificial Intelligence Help Prevent Online Sexual Abuse? |
| **Navn på veileder *:** | Matilda Dorotic |

| | | | |
|---|---|---|---|
| **Inneholder besvarelsen konfidensielt materiale?:** | Nei | **Kan besvarelsen offentliggjøres?:** | Ja |

### Gruppe

| | |
|---|---|
| **Gruppenavn:** | (Anonymisert) |
| **Gruppenummer:** | 206 |
| **Andre medlemmer i gruppen:** | |

Master Thesis

# - How Can Artificial Intelligence Help Prevent Online Sexual Abuse? -

Handed-in date:
28.06.2022

Campus:
BI Oslo

Examination code and name:
GRA1974 Master Thesis

Programme:
Master of Science in Business with a major in Marketing

Supervisor:
Matilda Dorotic

## Acknowledgements

We are grateful for our supervisor, Dr. Matilda Dorotic for helping us with our Master Theis and introducing us to preventing Online Child Sexual Exploitation and Abuse with help from technology like artificial intelligence. The elective course “GRA6811 Artificial Intelligence, Algorithms and Society” by Christian Fieseler and Eliane Bucher also inspired us to write about artificial intelligence. We appreciate the inspiration and learning this course provided us.

It has been an interesting experience studying such an important topic. We would like to thank our supervisor for always giving us constructive feedback and being available regardless of your workload. We are very thankful.

We would also like to thank everyone who participated in our study and helped us complete our research. Especially our friends and family, who have supported and encouraged us throughout the whole thesis.

Sincerely,
Charlotte Vaktel and Lizeth Brandsås

# Executive Summary

Due to the increasing problem of Online Child Sexual Exploitation and Abuse (OCSEA) over the past years, finding a way to prevent it has become more important. There has never been easier for perpetrators to contact children and hide their identity. The problem has increased to such an extent that governments around the world consider new laws that would have to breach privacy in order to prevent the spread of inappropriate materials through digital platforms. The firms and marketing will have to play a critical role in making this change. But the change is flooded with serious tensions and challenges on all sides: for firms, governments, and law enforcement. Marketing has a pivotal role to help explain and remedy some of those challenges. This thesis aims to investigate how people would react to being surveilled on social media platforms with the intention of preventing OCSEA. If consumer's privacy is to be severely breached, who should do the message screening: an artificially intelligent robot, or a human? The increasing health problem has contributed to the European Commission proposing a derogation that allows tech companies like SoMe platforms to derogate from the EU privacy framework in order to detect, report, and remove abuse material and prevent OCSEA. The United Nations Convention on the Rights of Children says that children under 18 years have the right to protection and the right to be heard (FN-sambandet, 2022), therefore, preventing crimes against children is important. Not only can it be harmful to the child physically, but also mentally. Several physical ailments are a consequence of such abuse. Fortunately, there are technologies like artificial intelligence (AI) that can help detect, report, and prevent this type of abuse. Even though we live in a technologically developing world, there is still a lot of scepticism towards AI. In general, research evidence indicates that humans are always preferred over AI in decision making that may deal with some type of a moral issue.

We also investigate ways in which new regulations in preventing online abuse could be achieved. We want to understand if people will be more accepting of an AI robot rather than a human surveilling their online activity and messages, and in which scenarios they allow it. Their preferred choice is measured by their trust in AI, fear of being misinterpreted by AI, and fear of being discriminated against. In addition, we investigated different factors like people's anxiety levels and privacy concerns when measuring the research question. To check if other variables affected people’s choice of conductor, we looked at several moderators such as “Age” and “Gender”. In addition, we checked for “The purpose of surveillance” to see what purposes would be accepted.  An extensive survey was conducted to map people’s preferred choices. In conclusion, our study found that most people prefer an AI robot over humans to surveil their online activity and messages. It also shows that people are more accepting of surveillance if it is for the betterment of society, rather than for commercial and advertising purposes.

**Table of Content**

## List of Abbreviations

AI – Artificial intelligence

CSA – Child Sexual Abuse

CSR – Corporate Social Responsibility

DV – Dependent variable

GDPR - General Data Protection Regulation

IV – Independent variable

SoMe – Social media

OCSEA - Online Child Sexual Exploitation and Abuse

## 1.0 Introduction

The United Nations Convention on the Rights of Children says that children under 18 years have the right to protection and the right to be heard (FN-sambandet, 2022). Despite these rights, there is an increasing worldwide problem where children are being maltreated (Sethi, et al., 2018). There has been a dramatic increase in reported cases in Norway related to online child sexual exploitation and abuse during the last few years (Oslo politidistrikt, 2022). According to the Terminology Guidelines for the Protection of Children from Sexual Exploitation and Abuse, OCSEA defines "all forms of sexual exploration, and abuse of a child carried out directly online or facilitated, in whole or in part by the digital environment" (Sylwander, Vervik, & Greijer, 2021, p. 6). Due to current technology, it has never been easier for adults to connect with children online (ECPAT Norway, 2021).

From 2015, there has been an increase in reported sexual offenses in Norway, where most cases involve social media. Especially during and in the aftermath of the Covid-19 pandemic, the reported cases of OCSEA have increased (Oslo politidistrikt, 2022). In Norway, 97% of children between the ages of 9 and 18 years have mobile phones, and 90% of the children in that group use social media platforms (Medietilsynet, 2020). Social media (SoMe) is defined as “interactive technologies and digital channels that allow people to communicate and share information, ideas, interests, and other forms of expression through virtual communities and networks” (Cambridge Dicitonary, n.d.).

The digital revolution has created a digital setting that makes it difficult to control people's identities and hence, enabled a new place where abuse of children can happen. The use of SoMe platforms on smartphones contributes to the loss of control of what children are doing; what they share and with whom (Aanerød & Mossige, 2018). SoMe platforms do not require any identity check to create an account, making it easy to hide and fake the real identity of individuals (Mæland, 2021). The abuse often occurs between children and adults or between children themselves on SoMe platforms or other digital platforms by sharing messages, photos, videos, etc (Frøyland et al. 2021). The focus of this thesis is the OCSEA between children and adults. OCSEA is connected to mental health difficulties like depression, anxiety, suicidal thoughts, and decreased self-esteem (Aanerød & Mossige, 2018). These increasing health problems have made the prevention of OCSEA one of the European Commission's top priorities after 2017 (Sunde & Sunde, 2021).

The European Commission’s new derogation allows tech companies like SoMe platforms to derogate from the EU privacy framework in order to detect, report, and remove abuse material and prevent OCSEA (Dorotic, 2021; Mildebrath, 2022). The reporting started as a voluntary action, but since the requirements were voluntary, many companies stopped reporting in fear of breaking the new privacy regulation that was forced at the end of 2020 (Bateman, 2022). It has forced the Commission to change its current strategy and include clear laws with conditions and safeguards protecting both the users of these platforms as well as the potential victims (European Commission, 2022).

In the beginning of 2022, the Commission proposed new EU legislation obligates the providers to detect, report, and remove materials connected to OCSEA on their platforms (European Commission, 2022). This proposal has met a lot of criticism and scepticism because of its threat to people's privacy. Opponents are afraid that if this new law should be accepted, we will live in a world of surveillance since these platforms/firms/government agencies can look into our personal messages and online activity (Hern, 2022). The problem with OCSEA is that it is challenging to discover and even more challenging to prevent. Different privacy regulations protect SoMe users, creating an ethical dilemma between privacy and crime prevention. However, new technologies can be helpful in the investigation and prevention of OCSEA.

Artificial intelligence (AI) is a set of algorithms and can be defined as “the theory and development of computer systems that are able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision making, and translation between languages” (Oxford Reference, 2020). It can be a digital machine’s ability to perform tasks commonly associated with intelligent beings (Copeland, 2022). There is an enormous potential to improve productivity by including AI technology to make decisions, recommendations, and predictions (Smith, 2020). Searle (1980) describes AI as “a powerful tool’’ that enables us to examine larger and more precise amounts of data than human brains could (Bechmann & Bowker, 2019; Searle, 1980). If companies use AI to conduct automated tasks like data collection and screening, it can save the companies a lot of time and effort (Bucklin et al., 1998).

Even though we see that AI can be a helpful tool for companies when making decisions, the question is in what situations it would work and if the SoMe users would approve it. To prevent OCSEA on SoMe platforms, the firms need to access the users' personal data. Personal data is described as "any information that relates to an identified or identifiable living individual" (European Commission, n.d.). Information about the user, including their private messages, pictures, and other shared content, may be of interest in a potential investigation. As a SoMe user, having your private messages looked into by another human can feel intrusive. AI can emerge as a non-judgmental third party when screening messages and reviewing data.

In some situations, AI may be perceived as less judgmental, making it less embarrassing if an AI-based robot looks into your personal messages rather than a human. Even though the technology is "smart", AI is not able to evaluate, judge or use the information in the same way a human could (Cole, 2020). When screening messages, AI will only look for specific words and sentences, without any intention of using it. In that way, the screening process may feel less intruding for the involved parts. However, the inclusion of AI provides a considerable responsibility for SoMe companies, and they must be conscious of how they use it.

Today, almost every company uses AI somehow and has adopted its benefits into their work. AI is often used in hiring processes, but one concern is that AI will be biased, creating dissatisfaction (Bertrand & Mullainathan, 2003). Over the years, people are getting used to the implementation of AI and algorithms, but that does not mean they trust it to make moral decisions. People are sceptical about AI making morally-relevant decisions because of the missing human mind, feelings, and potential biases. This scepticism is also known as AI aversion. (Bigman & Gray, 2018). As many people are sceptical of AI, the companies who wish to include it, face quite a challenge overcoming the aversion.

Moreover, some people fear that AI technology and robots will be able to outsmart humans in the future (Newman, 2017). Elon Musk, the founder of Tesla and SpaceX, has stated that AI is “potentially more dangerous than nukes” (Floridi, 2016, p. 1). AI aversion and resistance from the SoMe platform users may be an obstacle in the work of preventing OCSEA. Because of this scepticism, it is not clear if or in what situations the implementation of AI in SoMe decision making would work. Therefore, this study examines the SoMe users' reaction to AI decision making in crime prevention situations like OCSEA.

SoMe Companies have to take Corporate Social Responsibility and act in a manner that protects their stakeholders; users, society, the police, investors etc. Corporate Social Responsibility (CSR) is the standard regulation for companies of how they should behave. From a marketing point of view, it is crucial for the SoMe platforms to maintain their customers, which we refer to as “users” in this thesis. Marketing is driving the CSR activities and it is therefore the marketing function in the firm that makes firms realize the consequences of social impact. Hence, the marketing department, along with other departments such as the legal department, Know Your Customer (KYC) team, and the analysts, must find a proper way to tackle the new regulations. If the new derogation is accepted the firms marketing will play an important role in the prevention of OCSEA. Marketing is essential to get the user’s trust as well as tackling the social harms that come as a side effect of the derogation. Especially since SoMe companies collect information about their users’ online activities, the users must feel safe when using the platforms. User data is strictly regulated by laws like the General Data Protection Regulation (GDPR) and should only be used for the purposes described in the terms of privacy.

However, the user data provides a huge opportunity for AI-based customer surveillance, making it possible to monitor the shared content, provide evidence, and possibly predict crimes in advance. On the other side, surveillance of users’ online activity on SoMe platforms can jeopardize the relationship between the user and the platform. If the new legislation from the European Parliament is accepted, it can harm the users’ perception of the SoMe platforms. It will be a fine line between utilizing the information without making the users feel surveilled.

This thesis aims to understand if users will accept SoMe platforms collecting user data and if the users will be more accepting in situations of crime prevention rather than commercial purposes. Surveillance of online activity using algorithms are often used for commercial purposes and marketing. Little research is provided on the prevention of OCSEA, but due to the increasing problem and the purposed derogation, it is an upcoming field of research. In the light of the recent increase of criminal cases involving OCSEA, we find the topic of this study highly relevant. We will use different bodies of academic literature to understand how SoMe companies can contribute to preventing OCSEA. Based on this, the research question of this study is:

> *When firms have to breach privacy to prevent child abuse on social media platforms, would users be more accepting if message screening of their online activity is conducted by an automatic AI detection tool rather than humans?*

## 2.0 Literature review

This part will present the theoretical framework for this thesis, including definitions and existing literature that will form the basis of the study. The literature review is divided into three parts, the first one focusing on SoMe platforms and Norwegians' digital activity. The second part elaborates on artificial intelligence and how it can help firms with their decision making. The third and last part will address the SoMe platforms' role and responsibility, in addition to people's online privacy.

### 2.1 Social media platforms and people's digital activity

This part will start with an overview of Norwegians' activity on SoMe platforms. Followed up by an explanation of online child abuse, and how SoMe platforms can make it easier to get in touch with children. As well as a part of how legislation requires SoMe to help prevent child abuse.

### *2.1.1 Social media activity in Norway*

Due to the digital revolution, we have gotten new and easier ways to communicate and share photos, thoughts, and opinions. In Norway, 98% of the population have internet access at home, and 96% of the people in the age 9-79 years have their own phone (Statistisk sentralbyrå, 2022). This makes the possibilities for using SoMe platforms high. Norwegians are on the top of the world using SoMe platforms, with 8 out of 10 using SoMe daily (Christensen, 2021). SoMe platforms were initially made for interaction with friends and family but are also used by companies to reach out to users as potential customers. (Dollarhide, 2021).

Given the prosperity in Norway, larger parts of the population use SoMe platforms, including children and older people (Aanerød & Mossige, 2018). 78% of the Norwegian population over 18 years are using SoMe daily, according to Ipsos SoMe-tracker Q1 '22. Facebook is the most popular platform, with 82% of the Norwegian population as users, followed up by Instagram (67%) and Snapchat (66%) (see Appendix 1) (Ipsos SoMe team, 2022). The survey (N=6000) by Ipsos (2022) also shows that TikTok is the most growing SoMe platform in Norway, but still, only 23% of the Norwegian population over 18 years use it.

97% of the children between the ages of 9-18 years have mobile phones, and 90% of the children in that group are using SoMe platforms, according to a report by Medietilsynet (2020). The most common platforms for this age group are Snapchat (80%), TikTok and Instagram (both 65%), and Facebook (51%). (Medietilsynet, 2020) The report also addresses that the percentage increases significantly with age, mainly from 13-14 years. According to the Ipsos SoMe Tracker Q1 '22 (see Appendix 1), it is still Snapchat, followed by TikTok and Instagram, which are the most used SoMe platforms by children under 19 years in the first quarter of 2022.

Unfortunately, more than 3 out of 10 children who use SoMe regret something they have shared (Medietilsynet, 2020). Girls tend to regret the most, with almost half of the group regretting having shared something on SoMe. The report also found that 3 out of 10 in the age group 13–18-years have gotten sexual comments online. Most girls have gotten these comments, and 35% described it as disgusting, while 50% blocked the person who commented. In the age group 13-18, 42% of the participants had been asked to share a nude photo by a stranger online. 13% of these have gotten paid for sharing nude pictures. (Medietilsynet, 2020) Furthermore, 40% of the participants have received nude photos from strangers online. A common denominator in several questions in the research is that girls are more exposed to unpleasant incidents on SoMe than boys. According to Aanerød and Mossige (2018), teenage girls are most vulnerable to sexual assault, but boys are also exposed. Young people in the age range 12-18 spends a lot of time online and communicating to others on SoMe platforms. Their use of internet makes them available, also for those who want to establish sexual contact.

### *2.1.2 Online child abuse and grooming*

The increasing popularity of SoMe platforms in the last decades has created new ways to communicate with people that previously would be difficult to reach. Child maltreatment is a major public health problem, and is defined as "the psychical, sexual and/or emotional abuse and/or neglect of children under 18 years" (Sethi, et al., 2018, p. viii). Due to this new way of communicating, online child sexual abuse has become a bigger issue (Sunde & Sunde, 2021).

Child sexual abuse (CSA) is defined as the crime of harming a child physically, sexually, or emotionally (Oxford Learner's Dictionaries, n.d.). CSA can be divided into three different categories: penetrating abuse (including oral abuse), contact abuse (sexual touching), and noncontact abuse (exposure to filming, and other forms of sexual activity not involving physical contact (Gilbert et al. 2009; Kloppen et al., 2016). Self-reporting surveys among children and teenagers and reported CSA cases shows that the offenders are increasingly using the internet to find their victims (Kripos, 2019). OCSEA is a cybercrime where technology plays a role across a broad spectrum of activities connected to abuse or exploitation (Quayle, 2020). There are three types of OCSEA; the first one is "Live online child sexual exploration and abuse", which refers to only online abuse. The second is "Child sexual abuse material" (CSAM), which is photos, recorded videos, or other materials of sexual abuse. The last one is "Online and offline sexual abuse", which is a partly committed online but also have offline components. (Sylwander, Vervik, & Greijer, 2021)

From 2015, reported sexual offenses in Norway has increased, with the largest increase involving SoMe cases (Oslo politidistrikt, 2022). Illegal photos and films stand for 2/3 of all the tips, and criminal chatting between adults and children stands for 1/3 of the tips (Aanerød & Mossige, 2018). Reported rapes of children under 14 years have continued to increase, especially in the aftermath of the pandemic. In 2020 there was reported twice as many cases in Oslo Politidistrikt as in the pre-pandemic period from 2015 to 2019 (Oslo politidistrikt, 2022). In 2016, 5% of the reported rape-cases in Norway were started online (Ertzeid, 2021). From January 2015 to November 2020, there were 223 convicted cases of OCSEA in Norway, with 1336 victims (Sylwander, Vervik, & Greijer, 2021). Research shows that many of the convicted adults of OCSEA have interacted with many children simultaneously (Aanerød & Mossige, 2018; Sunde & Sunde, 2021). The police consider that online abuse is becoming more serious (Aanerød & Mossige, 2018). Unfortunately, the police and the government struggle to keep up with the technology in this area (ECPAT Norway, 2021; Sylwander, Vervik, & Greijer, 2021). Hence, something must be done. Firms in particular are called to action, and marketing and customer protection strategies are particularly relevant for these issues.

SoMe platforms make it easier for adults to get in contact with children. Grooming is defined as "the action by an adult who tries to become friends with children, particularly through the internet, with the intention to have a sexual relationship" (Oxford Learner's Dictionaries, n.d.). At least 2 out of 10 children under 18 years have experienced sexual contact online within the first year that the contact was iniated, according to the NOVA-rapport "Seksuelle overgrep mot barn og unge via digitale medier (2021)". Aanerød and Mossige's (2018) research on the Norwegian population from 2015-2017 indicates that OCSEA is a significant societal problem that needs more extensive research. Child abuse on SoMe has gotten more attention in the last few years, but there is still little research done in this field in Norway (Aanerød & Mossige, 2018; Ertzeid, 2021; Frøyland et al., 2021). We have tried looking through different databases like the international "Web of Science" and "Google Scholar", and the Norwegian databases "Idunn" and "Nasjonalbibliografien". It showed that little research is conducted on sexual abuse, especially in connection to SoMe in Norway. The 7 publications found have been published in recent times from 2016. The research in this field is mainly provided on behalf of the government, the police, or voluntary organizations such as ReddBarna.

Regarding the NOVA report (2021), 1 out of 10 youths has experienced inquiries from adults via SoMe. There are no statistics on the number of young people exposed to grooming because of the limited research on this field in Norway and Scandinavia. It is also difficult to measure grooming because it is defined differently. Common for the definitions is an illegal act by an adult trying to have a sexual relationship with children. In addition, there are many occasions where the child thinks they communicate with peers and never finds out that this is an adult. The three most common ways for groomers to hide their real identity when reaching out to children are pretending to be younger than their actual age, pretending to be a child or a teenager, and using fake pictures (Sunde & Sunde, 2021). Another common way to reach out to children is to lie about their gender. Lying about age and gender has been used in several Norwegian OCSEA cases (Bergen, 2014; Sunde & Sunde, 2021). Grooming can also appear in real life but is mainly associated with online activity. Studies have shown that men often perform grooming (Ertzeid, 2021; Frøyland et al. 2021; Sunde & Sunde, 2021; Aanerød & Mossige, 2018). These men are usually between the age of 25-45 and have a high technological competence which helps them fake their identities (Bergen, 2014; Frøyland et al. 2021; Sunde & Sunde, 2021). Snapchat is anonymous and encrypted, and therefore an easy way to connect with children. Chats and pictures are not saved, making it harder to find out who is behind an account. The number of reported cases on OCSEA is increasingly connected to Snapchat (Mæland, 2021).

Norwegian legislation lacks a requirement for SoMe platforms and technology companies to protect their users, and prevent and report OCSEA (Sylwander, Vervik, & Greijer, 2021). According to the report by Aanerød and Mossige (2018) Facebook and Movie Star uses technology to surveil their platforms for inappropriate materials. It is still difficult to avoid and prevent OCSEA on these platforms. SoMe platforms and other online companies have an increasing expectation of taking responsibility to detect, report, and remove materials connected to OCSEA.

### *2.1.3 Preventing online child abuse and grooming*

OCSEA is a crime that can have long-term consequences for the victims, and society, for example mental difficulties. Preventing this crime is a complex challenge that requires actions by the governments, as well as the firms where the crime can happen (Council of Europe, n.d.). Since 2017, preventing OCSEA has become one of the European Commission's top priorities (Sunde & Sunde, 2021). In line with the EU laws, the Norwegian government has ratified the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS 201) (The Norwegian Ministry of Children and Families, n.d.).

On the 2nd of August 2021, the ePrivacy derogation came into force, a temporary derogation from Directive 2002/58/EC (Aguilar, 2021; Bertuzzi, 2021). The derogation allows tech companies like SoMe platforms, telecommunication companies and similar digital service providers to derogate from the EU privacy framework. The derogation aims to prevent and remove OCSEA and connected online materials voluntarily for the companies. According to the commission, the strict privacy framework helps these firms use specific technologies to prevent and remove child abuse and related materials and, at the same time, protect people's privacy (Mildebrath, 2022).

After the new privacy regulations and GDPR was forced in 2020 many companies stopped reporting in fear of breaking these regulations and breach people's privacy (Bateman, 2022). Almost all (95%) of the reported cases of OCSEA received in 2020 came from only one company. The company's behaviour, and lack of reporting shows how deficient the current voluntary reporting has been. It has forced the Commission to make clear laws with conditions and safeguards protecting both the users of these platforms as well as the potential victims. (European Commission, 2022).

On the 11th of May 2022, the Commission proposed a new EU legislation that obligates the providers to detect, report, and remove materials connected to OCSEA on their platforms (European Commission, 2022). The proposed derogation has raised much concern and anxiety regarding people's privacy. People are especially concerned about mass surveillance and its threat to people's privacy. The opponents are afraid of a world of surveillance if the proposal is accepted since it allows SoMe platforms to look into our messages and online activity (Hern, 2022). The Commission argues that this proposal is highly needed because of the increasing confirmed reports of OCSEA with 64% more reports in 2021 compared to 2020. They also assure that the strict regulations will prevent mass surveillance. The detection can be done automatically and anonymously by technology like artificial intelligence, to minimize the impact on the user's privacy. Humans will only be involved in the process if there are concrete suspicions (European Commission, 2022).

## 2.2 Artificial intelligence and decision making

The following part will focus on how AI and algorithms can help prevent crimes of OCSEA on SoMe platforms. It will also describe why people are sceptical about including AI technology in decision making.

### *2.2.1 Artificial intelligence and prevention of OCSEA*

Crime prevention on SoMe platforms is becoming an increasing priority worldwide, but it requires advanced technology. For companies to handle and benefit from data collection, AI, automation, and advanced algorithmic systems are crucial (Bucklin et al., 1998). There has been a rapid increase in the integration of algorithms and autonomous machines into human society in the last decades. Today, people rely on different types of algorithmic advice, for example, Spotify's music recommendations, Netflix's movie recommendations, and LinkedIn's job suggestions. Algorithms can be described as a mathematical set of rules that specify how a group of data is behaving (O'Brien, 2022). Previously human-performed tasks can now be performed by algorithms and autonomous machines (Bigman et al., 2022). Another type of AI is Natural language processing (NLP), which makes decisions based on data obtained from human language (Russell & Norvig, 2016). Some benefits of algorithmic tools are that they are sometimes faster, cheaper, and make fewer errors than humans. Therefore, including technology enables more complex jobs to be completed at a higher speed.

Big data is another new phenomenon and has become one of the most shaking technological advancements, leading companies to invest large amounts in software development. Big data is defined as “data that contains greater variety, arriving in increasing volumes and with more velocity.” (Oracle, 2022) Companies collect and store large amounts of user data to get the advantages of the new technology both now and in the future (Alshura et al., 2018,). Such data provides companies with good opportunities for analysis, prognoses, and predictions due to the large datasets collected.

Crime prevention, especially prevention of OCSEA, is a growing research field (Sunde & Sunde, 2021). Two examples of AI in crime prevention are Sweetie 2.0 and AiBA (Author input Behavioural Analysis). They are both AI-based prevention methods against online OCSEA. Sweetie 2.0 is a research project with the aim of preventing webcam sex tourism. It uses both a chatbot and a virtual 10-year-old girl from the Philippines. (Sunde & Sunde, 2021; Terre des Hommess, 2022) After ten weeks of operation, Sweetie found one thousand potential offenders from 71 countries. Despite its success, Sweetie 2.0 has stopped its activity because of conflict with its ethical standards and codes of conduct (Terre des Hommes, 2022). AiBA is a Norwegian development that applies linguistic and behavioural patterns, like words and writing, to predict the gender and age of participants in an online conversation. The purpose is to warn children if they are talking to adults posing as a child. (Aaskervik, 2020; Sunde & Sunde, 2021).

Another example of how AI can be used for crime prevention is as a resource for community supervision officers. In that way, they can monitor offender behaviour and potentially stop offenders at risk of committing new crimes before it happens (Martin & Moore, 2020). The same method could apply to SoMe platforms to make the users behave and not contribute to the spread of hate comments, threats, and abuse online.

Based on this, we see that AI can be helpful in the prevention of OCSEA, but the question is how companies can sufficiently use the technology. According to the proposed law by the European Commission, AI should be used to maintain the users’ privacy when detecting potential crimes of OCSEA. Even though AI can be beneficial in preventing OCSEA, the question is if we can leave all the responsibility to AI when deciding if someone is potentially guilty of an OCSEA crime. It is still unclear in what situations AI's decision making will work and how helpful it will be. In particular, the open research question is how customers will react if the SoMe platform they use would conduct surveillance of their personal messages.

### *2.2.2 Making the right decision*

Making the right decisions can be difficult for humans, especially alone since they do not have the capability to take all relevant information into account (Frankenfield, 2021; Herbert, 1965). Luckily, new technology can help firms to get an overview of all information needed to make the best possible decision. How SoMe companies make their decision is essential for their reputation and relationship with their users.

#### *2.2.2.1 Decision making in companies*

Decision making in companies is a well-researched field defined as "the process employed by an organization that establishes its goals and the strategy to achieve these" (The Editors of Salem Press, 2016). It is seen as one of the most central processes in a company, and is required at all levels (Li, 2008). For the new derogation to work, the SoMe companies must understand how they can make the right decisions when preventing OCSEA on their platforms.

Decision making in companies is a well-researched field defined as "the process employed by an organization that establishes its goals and the strategy to achieve these" (The Editors of Salem Press, 2016). It is seen as one of the most central processes in a company, and is required at all levels (Li, 2008). The aim is to understand how SoMe companies can make the right decisions when preventing OCSEA on their platforms.

Only fifty years ago, the decision making process in firms were mostly based on human judgment. Managers trusted their gut feeling and human judgment, developed from experience. At this point, a relatively tiny bit of data was used to make decisions (Colson, 2019). A lot of research shows that higher-quality decisions result from incorporating a broad range of information into the decision making process (De Cremer et al., 2011).

Two of the classical theories in organizational decision making will be presented, to better understand why AI can be helpful for managers in the decision making process. One of the first and most well-known theories about decision making is The Classical Rational Decision Theory. The theory assumes that decisions shall rely on rationality and optimization to maximize the goal (Hoy, 2019; Li, 2008). This theory often describes the company's perspective including the goal to maximize their profit and vision. However, this theory is criticized for not being realistic to real-life decisions, due to the lack of other factors than rationality included in decision making (Herbert, 1965; Li, 2008).

Herbert Simon is one of the first to highlight the importance of decisions in firms with his model "Decision Making Theory" presented in his book, Administrative Behaviour (1947). Simon's theory is based on The Classic Rational Decision Theory and wishes to understand how decisions are made in firms (Simonsen, 1994). This theory takes the human mind into account and considers the psychological aspects that The Classic Decision theory overlook (Harappa, 2021; Li, 2008). It is based on a more realistic view of the human brain, assuming that humans seek to use the available information to make a satisfactory decision, or one that is "good enough" (Frankenfield, 2021). It is impossible for humans to obtain all information to make a fully rational decision (Frankenfield, 2021; Herbert, 1965).

It has become essential for modern firms to be able to deal with ethical issues due to their role in society and expectations from their users and customers. Jones (1991) defines ethical decision as "a decision that is both legal and morally acceptable to the larger community" (Jones, 1991, p. 367). Ethical decisions deals with moral issues: "A moral issue is defined as individual actions, when freely performed, may harm or benefit others" (Jones, 1991, p. 367). Morality is often believed to require a fully human mind that can both feel and think in more psychological terms (Bigman & Gray, 2018). Research shows that stress is connected to managers ethical decision making, and often has negative effect on decisions (Selart & Johansen, 2011). Managers tend to have busy, demanding, and stressful work schedules (Ganster, 2005; Hambrick et al., 2005). Therefore, it is important to understand how stress influence decision making to ensure the best possible decision for all the involved parts. New technology can be beneficial because it can help firms to make the best ethical decision by reducing stress and include all the relevant information.

#### *2.2.2.2. Decision making including AI*

After the digital revolution, information, and analysis methods such as the Internet became common in decision making (Citeron, 2011). Already in the 1950s, Herbert Simon and his colleague Allen Newell attempted to simulate human decision making on machines. They also tried to include AI in the decision making process. (Frankenfield, 2021; Gugerty, 2006) According to Simon, machines would be capable of doing any work a man could do (UBS, n.d.).

One of the most outstanding scientists of the twentieth century was Alan Turing. Turing introduced the "Turing Test" with the purpose of determining if the machine was capable of thinking. The test used both a computer and a human, using an interrogator to decide which of the two contestants was a human or a machine. The test was passed if the computer were able to convince a human that it was a real person (Russell & Norvig, 2016). The Turing Test will remain important in the development of AI and has brought real relevance to future generations of people living in a world where the cognitive capacities of machines will be more extensive (French, 2000).

One of the advantages by including AI is its ability to notice patterns in large datasets, way beyond human perception. With the help from the technology, we now can capture huge volumes of data like every transaction and/or every user gesture that can help us to make good decisions. (Colson, 2019). AI has increased the amount of accessible information and decreased the amount of time collected (The Editors of Salem Press, 2016). Today, many jobs make use of algorithms as an aid in making decisions (Prahl & Swol, 2017). Different types of algorithms, for example, chatbots, forecast systems, content-creators, and speech recognition, are increasingly becoming a part of companies. The mentioned types are often used as tools for various decision making processes (Prahl & Swol, 2017).

Automation is highly appreciated when considering efficiency or effectiveness (Bucklin et al., 1998).

"Experts in various professions—including medicine, psychological counselling, human resource management, banking, science, transportation, public administration, and legal counselling —increasingly rely on the guidance of AI-based algorithms when making important decisions." (Shrestha et al., 2019). The new way of making decisions is challenging the old "human-based" decision making process, without any guarantee of being the preferred way by the company's users. However, other professions relying more heavily on AI decision making support the possibility of AI-based decision making in crime prevention on SoMe platforms. To prevent crimes on SoMe platforms, the companies need to decide what is perceived as illegal, harmful, and unacceptable content, and then choose how to handle it. Big data provides the SoMe companies data that can be used in decision making in cases of OCSEA if they have the right tools to analyse it. The more user data a company holds, the more responsibility will be expected in making the morally correct decisions to protect the users.

One benefit of using AI in message screening is its ability to behave neutral, only looking for specific codes and warnings that can indicate a red flag. A machine may appear to understand a language but does not obtain a real understanding. Hence the "Turing Test". AI do not know if it is looking at a picture of a dog or a bread, and it do not even matter because AI is only required to recognise and not draw conclusions. Machines use syntactic rules to manipulate symbol strings but have no understanding of meaning or semantics. (Cole, 2020) This concept is building on "The Chinese Room" which was first published by Searle (1980). The concept of "The Chinese Room" is to simulate an understanding of the Chinese language by manipulating symbols and numerals just as a computer does, making it appears as if one speaks Chinese (Searle, 1980). AI will not be judgmental of what type of person you are or care about anything that you have posted. Therefore, AI can make message screening less embarrassing and easier to implement. However, a human is often required to ensure the quality of suggestions made by AI.

Advanced technology provides SoMe platforms an insight into the user's online behaviour. However, SoMe companies need to access the users' personal information to prevent OCSEA. The question is whether the users will accept this type of surveillance or resist it. The ethical aspect of using machines and AI robots in decision making can present challenges in terms of lacking the psychological elements. As the classical theory on decision making shows, the human mind is not capable of making perfect decisions and will never have the skills of a machine. The question is if a machine can make sufficient decisions without the human qualities of feelings and rationality. We will now explain why people are sceptical of companies' data collection and why some people become AI averse.

### *2.2.3 Why people are sceptical of AI*

People are getting used to the implementation of AI and algorithms, but that does not mean they trust it to make moral decisions. The use of AI in SoMe platform's decision making can make crime investigation feel less embarrassing when screening messages and looking through content. Despite people's awareness of algorithms, the most recent research shows that people are still sceptical of algorithms (Bigman & Gray, 2018).

The rise of algorithms also raised concerns about people's distrust of machines' decision making capacities (Bigman et al., 2022). Bigman and Gray (2018) in six studies found that people are averse to machines making morally-relevant driving, legal, medical, and military decisions and that the reason for this aversion is due to the perception that machines can neither fully think nor feel. Algorithm aversion can make people choose a human forecaster rather than evidence-based algorithms, even though research shows that algorithms can make future predictions more accurately than a human (Dietvorst et al., 2014). Since machines seem to lack a human mind, one can say that they may also seem ineligible for a machine to make moral decisions. As Bigman's studies revealed, preliminary evidence shows that people are averse to having machines make moral decisions. (Bigman & Gray, 2018)

*However, these findings were always in domains in which people decided about morality that did not directly impact them (deciding in hypothetical examples, like giving parole to prisoners, etc.). The research question pertains to whether the same feeling would persist if people needed to trade off a potentially embarrassing event for them or a privacy-intruding situation, like scanning private messages for inappropriate materials. In this context, would the potential embarrassment of being surveilled by humans overcome the algorithmic aversion?*

Some research explains that mistrust in algorithms is due to humans thinking they are the only ones capable of making perfect forecasts (Robyn, 1979). People sometimes rely more heavily on advice based on intuition rather than on algorithms (Onkal, et al., 2009). The aversion can be caused by people expecting an algorithmic system to work flawlessly, without unexpected errors in the algorithmic model (Madhavan et al., 2006). Algorithm aversion can become an obstacle when using data for crime prevention if the users do not accept it. Even though machines have good computational capacities, the ability to feel authentic emotion is still lacking and may cause a problem when including AI in decision making. One surprising finding was that people are averse to machines making moral decisions, even when they make the same decision, with the same outcome as a human (Bigman & Gray, 2018).

### *Biases and discrimination*

AI systems' predictive abilities are undeniable, but still, many people are uncomfortable with humanity's growing reliance on algorithms. Most people look at algorithms with a combination of fear and loathing, due to the inaccurate judgments that codes and number crunching can make (Bambauer & Risch, 2021). It started as simple and uncomplicated algorithms, but the technology is constantly developing. Today, algorithms can perform self-learning skills based on experience (Castelo et al., 2019).

Furthermore, AI raises the possibility of systematic discrimination being perpetrated by algorithms (Bigman et al., 2022). In a widely-cited NBER article, the experiments included resumes with "white-sounding" and "black-sounding" names. The investigation showed that the recruiters were more likely to select the white-sounding group (Bertrand & Mullainathan, 2003). The inclusion of AI presents risks involving the perpetuation of existing socioeconomic disparities potential, in addition to discriminatory or unfair outcomes (Smith, 2020).

In an attempt to overcome human biases in decision making, algorithms are a possible solution. Unfortunately, as with humans, algorithms sometimes make persistently biased decisions, discriminating against people based on their gender or race, for example (Bigman et al., 2022). Machine-learning algorithms discriminate because they are trained on biased datasets, thus reflecting existing social inequities. Unfortunately, unbiased decision making is unfulfilled since both AI and humans discriminate in many cases. (Bigman et al., 2022) However, when screening messages and going through data, AI can make it less embarrassing for the involved parts. The positive side of AI lacking feelings and emotions is that it may be perceived as less judgmental and personal in an investigation, leading towards AI appreciation.

### *AI Appreciation*

As a large part of academic theory on algorithmic decision making shows, people often prefer human judgment because they do not fully rely on decisions made by algorithms (Dietvorst et al., 2014). One interesting aspect is in what situations people prefer and accepts algorithmic decision making without behaving AI averse. Academic literature has found that people will not always be averse to algorithms, called algorithm appreciation (Logg et al., 2019).

Studies in this field, for example, Logg et al. (2019) revealed that decision makers appreciate algorithmic support instead of algorithm aversion as they adjust more towards estimates of algorithms than estimates of human agents (Jussupow & Benbasat, 2020). Several research articles have shown that algorithms can outperform the advice of humans (Dietvorst et al.,2014; Yeomans et al., 2019). In an article from Bambauer and Risch (2021) they tested if people prefer having a human versus an algorithm decide on an issue. The results showed some rejection of algorithmic decision making, especially in cases of a criminal trial. However, people opt for algorithms more often than one would imagen, based on factors like which costs less, which makes the fewest mistakes, and which is fastest. It shows that AI is more preferred in everyday situations, but when the stakes increase, people prefer a human to be involved. (Bambauer & Risch, 2021)

*Implementation of AI*

Because of a general scepticism towards algorithms, some academic literature provides approaches explaining how to decrease algorithm aversion, by training humans in for example leading positions to become more used to algorithms (Burton et al., 2019). The identified phenomenon of algorithm aversion creates a new research stream in psychology and management, investigating how user interaction with human agents differs from their interaction with algorithms (Jussupow & Benbasat, 2020). We will not look further into this field, as the relevant research will be restricted to the user's reaction to decision making that includes AI and not the interaction between algorithms and humans. As this part shows, the findings on algorithm aversion and algorithm appreciation are contradictory.

The inclusion of AI in decision making can be challenging in handling biases and making the users trust AI. SoMe companies must be aware of the potential resistance and work towards a common understanding that can benefit society and the users' safety online. The attention of different Governments and media awareness regarding this issue can help make the increasing problem of OCSEA understood and stopped. Another essential problem is the conflict between using user data for crime prevention and how this affects the users' privacy and the companies' reputation.

## 2.3 Social media platform's role and responsibility

To understand and answer the research question, the following part will explain digital privacy, and why surveillance of people's online activity can create a privacy problem. It is important to include privacy to get an overall understanding of the trade-off between using personal data for crime prevention versus protecting the user's privacy. It will also explain how the platform's responsibility and decision making can affect and harm the brand.

### *2.3.1 People's digital privacy*

Technology enables companies to record detailed data of any user transaction or behaviour, which gives them a major advantage: making it possible to deal with their users as individuals (Culnan & Armstrong, 1999). SoMe companies have experienced exponential user-growth since they entered the market, which has provided responsibility in terms of protecting their users (Sushama et al., 2021). As a result of extensive user bases, SoMe platforms store loads of personal and sensitive information that is only meant for the eyes of a specific receiver. Since personal data is including all the information that can be linked to a person in any kind of way, the companies must be conscious of how they protect the data (EUR-Lex, 2016;European Commission, n.d.).

Because of the big databases of stored personal data, concerns regarding privacy and personal data become an issue. When joining a SoMe platform, the users let the platforms access a lot of information about themselves and their online activity. For many people, it seems like a little reward to offer personal data to get the opportunity to easily communicate and share information with others (Zhang & Sun, 2010). Empirical evidence shows that people are willing to trade their personal information for relatively little reward. For example, one study by Carrascal et al. (2013) found that internet users were willing to sell their browsing logs for about 7 euros.

Another online study with approximately 4000 participants investigated different factors that influenced participant's choice of humans versus algorithms. In the presented scenario the decision maker needed to access the user's private information, and the results showed that privacy was not a meaningful factor. In addition, the participants were indifferent whether a human or an algorithm did the investigation (Bambauer & Risch, 2021). On the other hand, research has shown that personal concerns about privacy are a growing concern, and one of the primary obstacles to joining SoMe platforms or especially in electronic commerce (Gurung & Raja, 2016). The use of personal data as an organizational resource can create both positive and negative outcomes for the company and its users, based on how the information is used (Culnan & Armstrong, 1999).

One of the problems with these expectations seen from the SoMe platform's view is the trade-off between their users' privacy and their companies' social responsibility. If the platform does anything wrong, it can hurt the users "customer experiences" by breaching the privacy and trust of its users. (Dorotic, 2021) Therefore, a lot of SoMe platforms find this trade-off challenging. However, with the proposed legislation by the European Commission, this can and will change. SoMe platforms like Facebook, Movie Star Planet and Slettmeg.no all agree on the need for more knowledge regarding OCSEA and how to prevent it (Aanerød & Mossige, 2018).

### *2.3.2 Surveillance of personal data*

Customer surveillance is defined as the acquisition, usage, and storage of the user's personal data (Plangger & Montecchi, 2020). Surveillance gives companies the ability to surveil users, which can provide useful information in an investigation of a potential OCSEA situation. The use of personal data for this purpose may raise a threat to people's privacy, making them feel invaded or monitored by the companies.

Furthermore, this surveillance increases the users' data vulnerability, or unwanted use of their personal data (Martin et al., 2017). A study by Culnan and Armstrong (1999) looks into how users react to customer surveillance. The study proposes privacy calculus to clarify how consumers rational weigh the benefits and cost of disclosing personal data. Users are willing to disclose personal data and let companies make customer profiles for marketing if their concerns about privacy are addressed by fair procedures (Culnan & Armstrong, 1999). Culnan and Armstrong's ideas have dominated the privacy literature (Plangger & Montecchi, 2020). Users often claim to be concerned about their personal data but do very little to nothing to protect it. This disparity between individuals' information privacy attitude and their actual behaviour online is called the "privacy paradox" (Barth & De Jong, 2017; Kokolakis, 2017). There is a lot of empirical research trying to explain the paradox. One theory is the model of rational choice by H. A. Simon, which also explains individuals making decisions.

Content moderation and surveillance has been a prominent concern as serious issues have arisen on these platforms like revenge porn, child abuse, hate and discrimination (Grygiel & Brown, 2019). As we mentioned at the beginning of this thesis, the European Commission recently changed its plans to prevent online OCSEA, forcing SoMe platforms to report and remove possible harmful materials. This allows SoMe platforms to surveil and monitor their users' online activity and encrypted content (Tidey, 2022). Many big SoMe companies already do content moderation and surveillance on their platforms (Grygiel & Brown, 2019), which can be beneficial when preventing crimes like OCSEA in the future. Surveillance of the SoMe platform user's activity provides a lot of opportunities. However, the increasing concern for protecting the user's privacy is an important factor in not harming the SoMe platform's brand and reputation. Therefore, a trade-off occurs between a company's social responsibility and the relationship and satisfaction of their users.

### *2.3.3 CSR and brand harm*

Since the crimes of OCSEA is an increasing problem, the SoMe platforms are expected to take responsibility to prevent harmful activities on their platforms. It is essential to look at companies' responsibilities and how moral decisions can harm their reputation. Firms marketing is essential for keeping a good reputation and is closely connected to their social responsibility.

#### *CRS*

One challenge in preventing crimes on SoMe platforms is that there is no international governance for the internet, making it challenging when the platforms are global (Grygiel & Brown, 2019). Thus, globalization can make it challenging to create international laws that govern SoMe platforms. Given this challenge, studies argue that increased CSR is a key to a safer internet since it goes directly to the source; the companies (Grygiel & Brown, 2019).

Companies that operate online are increasingly expected to take responsible actions to detect, report, and remove abusive material and prevent cybercrime (Dorotic, 2021). They are responsible to "embodying those standards, norms or expectations that reflect a concern for what consumers, employees, shareholders, and the community regard as fair, just, or in keeping with the respect or protection of stakeholders' moral rights." (Carroll & Schwartz, 2011). Abuse and exploitation of children is an illegal activity and a breach of Children's Rights that should not happen. A platform where such behaviour can happen is expected to take responsibility for preventing OCSEA and show social responsibility. Preventing it will benefit the SoMe companies' stakeholders, the users, and society. The question is if CSR and the risk of brand harm will make firms less likely to engage in surveillance of private messages to prevent OCSEA.

The concept CSR is about the firm community's concern for society. CSR has existed for centuries, but it is most common in developed countries. Its long existence has led to an extensive body of literature on CSR. (Carroll, 1999) In the 1960s Keith Davis posed two intriguing questions about CSR which are the following: "What does the businessperson owe society?" and "Can business afford to ignore its social responsibilities?" (Carroll & Schwartz, 2011, p. 503). These questions started an important debate, bringing focus to the importance of protecting society. In this thesis, society is the perception of children exposed to OCSEA because it has become a societal problem. Regarding what CSR stands for, the prevention of OCSEA should be the foundation of SoMe companies' responsibilities.

It can be difficult for firms to know how to perform CSR and define their areas of responsibility. The social responsibility of companies encompasses the economic, legal, ethical, and discretionary expectations that society has of companies at a given point in time (Carroll & Schwartz, 2011). Studies on CSR in the 1970s started to define CSR more specifically. Later, the attention shifted increasingly to measuring initiatives and theoretical developments. CSR's ethical and legal perspectives are the most important in this study. The ethical domain includes activities based on their adherence to ethical or moral standards or principles (Carroll & Schwartz, 2011). To stop crimes on SoMe platforms, the companies need to engage in the work of preventing OCSEA, which they are encouraged to when ensuring the betterment of society. However, the prevention of OCSEA may harm the SoMe platforms' relationship with its users.

### *Brand harm*

Over the past decades, a growing number of ethical scandals have been discovered, which have raised a concern about unethical and irresponsible behaviour in companies (De Cremer et al., 2011; Hitt & Collins, 2007). To achieve a good reputation lies on the firm’s management, as they should make and implement decisions that result in the firm’s goals (Abubakar et al., 2019). The proposed derogation from 2022 allows SoMe platforms to use AI to surveil and report people’s messages and online activity. The derogation enables the prevention of OCSEA, but at the same time, the SoMe platforms get a higher responsibility to detect and report crimes.

As discussed, algorithms can help protect the users' privacy, but on the other hand, biases can create an issue for the SoMe platforms' relationship with its users. When algorithms go through and scan chats, they may falsely accuse users because of potential biases. For an AI robot, it can be challenging to spot what is the right and wrong activity online. Some examples of characteristics that can create these biases are gender, age, race, or other characteristics. (Dorotic, 2021) If the algorithms fail, they can cause harm to the company and its users, leading to brand harm crises (Srinivasan & Sarial-Abi, 2021). Falsely accusing a user of a crime can destroy the relationship between the users and the SoMe platform and ruin the platform's reputation. If companies use algorithms to detect crimes, the technology must consider potential errors and wrong interpretations to avoid brand harm.

It has been voluntary to report and remove possible occasions OCSEA. Still, it is no guarantee that platforms report or remove it, even if the content is offensive and harmful. Some companies are resistant because they are afraid of harming themselves (Srinivasan & Sarial-Abi, 2021). The new degenerations and laws can affect the SoMe companies' decisions and destroy the platform's reputation. Therefore, it can be hard to detect groomers and potential situations of OCSEA by surveillance.

The concept of CSR has been and will remain an essential part of companies' language and practice. The reason is that it is continually consistent with what the public expects of the company community today (Carroll, 1999). Equal trust between the users and the SoMe platform is essential for using algorithms in crime prevention. SoMe platforms have a unique opportunity to prevent OCSEA, which should be a priority for their CSR in the future. However, they must consider the potential brand harm it may cause.

### *2.4 Preventing OCSEA vs protecting users' privacy*

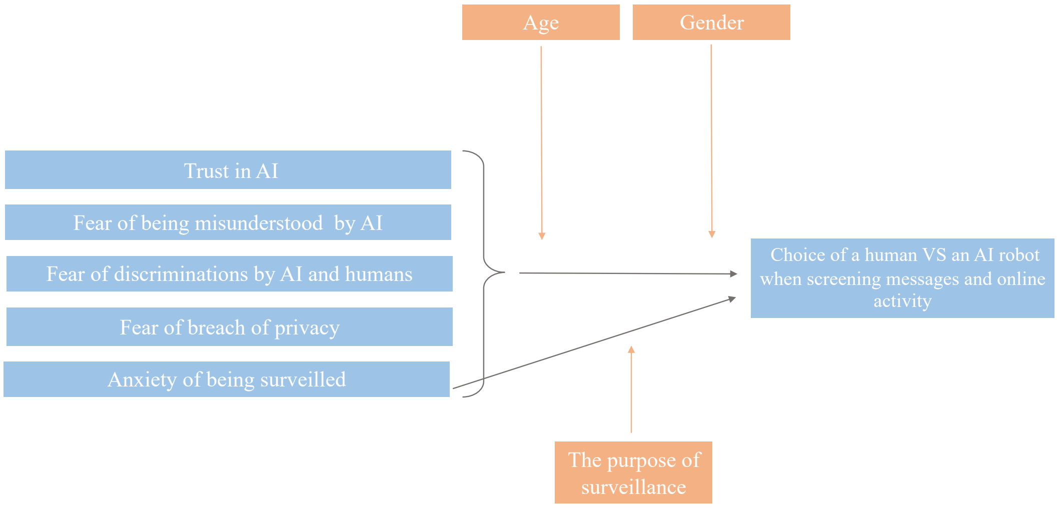
To sum up the literature, OCSEA is an increasing worldwide problem that needs to be solved. With the new proposed derogation from 2022 in mind, the SoMe companies face a challenge due to scepticism from the users in the implementation of the new rules. A lot of research supports the difficulties of the implementation of AI in various decision making processes. The literature also shows that people have little trust in SoMe companies due to scandals involving censoring information and monitoring their users. Thus, it can be challenging for SoMe companies to detect potential OCSEA crimes with respect to the harm it may cause the relationship with the users, users' trust, and brand reputation. At the same time, not acting can cause the same harms. Another issue is the users' privacy, and how SoMe platforms can protect both their users' privacy and the betterment of society. How they choose to proceed can impact the SoMe company's brand and reputation. However, the potential benefits of the new derogation are worth fighting for.

## 3.0 Conceptual framework

Based on the literature review, previous research has mainly focused on mapping how, where, and by who OCSEA is conducted, in addition to the number of exposed children. However, there has been little research regarding how to prevent OCSEA. Since OCSEA is an increasing issue worldwide, new laws to prevent it may include controversial and unethical methods. The proposed derogation by The European Commission has raised concerns regarding people's privacy online. That is why this thesis will try to understand people's acceptance of surveillance, both in general and in cases of crime prevention.

Since this thesis is based on The European Commission's proposed derogation from 2022, it aims to understand if the implementation of the new regulations is accepted by the society in Norway. We will focus on who would be the preferred conductor for message screening etc., a human or an AI robot. We wonder if there will be any clear preferences and if different scenarios will affect people's choices. Therefore, our dependent variable (DV) is the choice between a human and an AI robot when screening people's online activity and messages. The independent variables (IVs) will test people's trust in AI, their fear of being misunderstood by AI, people's anxiety levels, peoples fear of breach of privacy and fear of being discriminated by an AI robot or a human. To explain our conceptual model and illustrate our research, we made the following model (see Figure 1). The model shows the relationship between the variables and how the moderators affect the DV and the IVs. We want to test if different moderators can affect the participant's choice of a preferred conductor of message screening and surveillance of online activity.

Figure 1. *The conceptual model*


Age
Gender
Trust in AI
Fear of being misunderstood by AI
Fear of discriminations by AI and humans
Fear of breach of privacy
Anxiety of being surveilled
Choice of a human VS an AI robot when screening messages and online activity
The purpose of surveillance

*The moderators "Age" and "Gender" can affect all IVs, and the moderator "The purpose of surveillance" can affect the IV "Anxiety of being surveilled"*

### 3.1 Our independent variables

In this study, we have chosen five IVs that are expected to correlate with the DV. We believe these five variables are essential when answering the research question. Therefore, we will check the correlation between people's preferences and the IVs when choosing between humans and AI robots. It will be important to ensure that our chosen IVs are not mutually highly correlated to avoid strong multicollinearity issues. We will now further elaborate the chosen IVs and explain why we find then relevant.

### *Trust in AI*

People's trust in AI is essential to understanding how they will choose between a human or an AI robot. If people do not trust AI technology, it would be understandable if they prefer a human over an AI robot. As the literature review shows, people are often sceptical of AI making moral decision and lacking feelings, as well as potential biases (Bigman & Gray, 2018). In some cases where AI-technology has been used for decision making, potential selective biases have occurred (Bertrand & Mullainathan, 2003). Based on our literature review, people's lack of knowledge and general scepticism towards AI technology can impact the choice between humans and AI robots. We assume that the issue regarding trust in AI technology will affect our DV, and that our analysis will confirm this assumption. This IV was investigated by the question “Do you trust AI?”, measured by a 5-point Likert scale ranging from “Definitely not” to “Definitely yes”.

### *Fear of being misunderstood by AI*

Misinterpretation is a possible problem when communicating or interpreting information. It is not only humans that can misinterpret situations, AI technology can also have the same disability. The problem occurs when people expect algorithmic systems to work perfectly, which literature explains that it will not (Madhavan et al. 2006). On the other hand, the literature review shows that many people fear that AI will misinterpret situations, due to the inaccurate judgments that codes and numbers can make. We expect a correlation between the fear of misinterpretation and the preferred choice of the conductor in a potential OCSEA investigation. One reason for the correlation can be the fear of being misunderstood or falsely accused of committing a crime. If the fear of being charged with a crime is stronger than what contributes to preventing it, it can be challenging to implement AI into crime prevention. It will be interesting to see if the fear of being misunderstood will drive the participants towards choosing AI or a human conductor. To investigate this IV, the scenario used was: “Imagen a scenario where an AI robot is screening the messages between you and a friend, for an investigation of a case regarding child abuse. How anxious would you be about being misinterpreted by the AI robot? (knowing that you are innocent)”. We measured it by a 5-point Likert scale ranging from “Extremely anxious” to “Extremely unanxious”.

*Fear of discrimination by AI and humans*

From the literature review we see that many people assume AI technology to be flawless, even though it misses the human mind and feelings. Often, people do not imagine AI robots having the ability to discriminate. Unfortunately, including AI in the decision making process risks discriminatory or unfair outcomes (Smith, 2020). There have been occasions where AI has been discriminating, as in the example of white- and black-sounding names (Bertrand & Mullainathan, 2003). Still, we assume that people find humans more discriminatory than AI robots. People's precautions and knowledge of AI robots will probably affect their preferred choice of conductor. Thus, we think that the fear for being discriminated against is an essential variable to test in our study. The investigate people’s fear of discrimination by AI, this question “Do you think AI robots or humans are the least discriminatory?” was used. To measure this IV, we coded it as a dummy variable with two options (1 = AI robot, 0 = human). We see that the IVs “fear of being misunderstood” and “fear of being discriminated against” are both build on fear, but as it is perceived differently, we find it relevant to separate them.

*Fear of breach of privacy*

Since one of the biggest concerns of the proposed derogation is the threat to people's privacy, we find it important to include it. Research shows that people tend to care about their privacy. However, sometimes their actions show something else, creating the mentioned privacy paradox. On the other hand, research has shown that privacy concerns are growing among internet users. Privacy concerns may affect people's acceptance of surveillance, even if the purpose of the surveillance is the betterment of society by preventing OCSEA. For example, if you are concerned about your privacy online, it can affect your choice between AI or a human conductor. Based on earlier research, we assume there will be a correlation between privacy concerns and their preferred choice. However, this correlation might not be as strong as the correlation between the other variables. This IV was measured by using a 7-point Likert scale, ranging from “Low degree” to “High degree”. In the survey we used the statement “The following questions are about the personal feeling of privacy: - To which degree do you feel that social media companies protect your privacy?”

### *The anxiety of being surveilled*

Another concern regarding the derogation is the anxiety of being surveilled. Therefore, we believe that people’s anxiety level can impact our DV. Most people would probably prefer not to be surveilled, due to privacy concerns. In particular, some opponents fear living in a world of mass surveillance. We want to check if peoples fear of being surveilled is what influence them to choose either AI or a human. We expect to find a link between the fear of being surveilled and the preference of an AI conductor, since AI is perceived to be less judgemental. However, if the fear of being surveilled is too strong, one will probably wish to not be surveilled at all, by neither of the conductors. This IV was investigated by the assumption “Using social media platforms has made me feel anxious and afraid of being surveilled”. It was measured by a 5-point Likert scale ranging from “Strongly agree” to “Strongly disagree”.

An overview of how the variables were measured can be found in Table 1, below. The complete survey can be found in Appendix 2.

Table 1. *Overview of Variables and Measurements*

| Variable | | Scale | Measurement | Mean | Std.dev |
|---|---|---|---|---|---|
| 1 | Trust in AI | 1-5<br>(Definitely not – definitely yes) | *Do your trust AI?* | 3.09 | .0771 |
| 2 | Fear of being misunderstood by AI | 1-5<br>(Extremely anxious – extremely unanxious) | *Imagine a scenario where an AI robot is screening the messages between you and a friend, for an investigation of a case regarding child abuse. How anxious would you be about being misinterpreted by the AI robot? (knowing that you are innocent)* | 3.11 | 1.138 |
| 3 | Fear of discrimination by AI and humans | AI robot or human<br>(1= AI robot, 0= human) | *Do you think AI robots or humans are the least discriminatory?* | .7593 | .42953 |
| 4 | Fear of breach of privacy | 1-7<br>(low – high degree) | *To which degree do you feel that social media companies protect your privacy?* | 3.42 | 1.242 |
| 5 | Anxiety of being surveilled | 1-5<br>(Strongly agree – strongly disagree) | *Using social media platforms has made me feel anxious and afraid of being surveilled.* | 2.73 | 1.077 |

### 3.2 The moderators

Our model captures three moderators that alter the effect that our IV have on our DV. Since our study will focus on the choice of people's preferences when being surveilled online, it is essential to look at people's "Age" and "Gender" to better understand people's choice of conductor. These two moderators affect all five IV, and therefore, we will look at the effect on each IV. Our third moderator "The purpose of surveillance" will only check the impact on the IV "The anxiety of being surveilled".

#### *Age*

We have chosen to use "Age" as one of our moderators because people's age often has a connection with their technical skills and knowledge. Age can also determine people's knowledge of using the internet and acceptance of new technology, affecting their choice of conductor. Older people can be more resistant to new technology than the younger segment, but it does not mean they always are. Age is an individual factor that may or may not have a collective impact. However, we want to examine if age can create trends or parallels that affect people's choice of humans versus AI robots.

Age also affects how people use SoMe platforms and how often they use them, which can correlate with people's knowledge and acceptance of new technology. A survey on the Norwegian population in 2019 about time spent on SoMe platforms, showed that Generation Z spent almost five hours daily, while older people (above 72) spent nearly one hour ( Statista Research Department, 2022). Therefore, it is essential to check the moderator effect of people's age on the different IVs and how it can affect their preferred choice. For this moderator we used the question "How old are you?". We divided the population into eight different age groups ranging from 18 to 85+ years. Ratio scale was used to code the groups from 1-8 (see Table 2).

#### *Gender*

We will also investigate how gender can affect our IV. Research has found gender differences in time spent online and why they use SoMe platforms (Muscanell & Guadagno, 2021). Women are using more daily time on SoMe platforms than men according to Statista (Ceci, 2021). Because of different knowledge-levels due to time spent on SoMe, it will be essential to look at gender differences. Gender differences can therefore affect the IV, and thus it is essential to include them. For this moderator we used the question "How do you describe yourself?" and coded the data as a dummy (1 = Male, 0 = Female).

*The purpose of surveillance*

We want to check if the purpose of surveillance will have effect on people's choice of conductor, and which purposes people accept surveillance for. Therefore, we divided this moderator into two parts. As the literature showed, people are not very accepting of AI making moral decisions. However, people would probably not like to be surveilled by other humans because it can feel embarrassing and intrusive to their privacy. Different situations can be obtained differently, which made us want to understand if there are any differences in the choice of conductor if the aim of the surveillance is for commercial purposes or to prevent crime. Our impression based on the literature review, is that people's fear of surveillance and privacy concerns will surpass their trust in technology. Therefore, most people will choose an AI robot as their preferred conductor. We expect to see a connection between the purpose of the surveillance and the participants preferred choice of conductor. To answer this part of the "Purpose" we used a 5-point Likert scale ranging from "Strongly agree" to "Strongly disagree" on the assumption "I would allow social media platforms to look through my online activity if it is for the betterment of society (E.g. To avoid crime, child abuse, hate or discrimination)." (see complete survey in Appendix 2).

To check in what situations people will be accepting surveillance, and if different purposes will be accepted. The answer to this question will indicate if people are willing to extend further than they normally would, to accept surveillance based on the cause. Meaning that a person would to a larger extent agree to surveillance if it is for the betterment of society rather than for marketing or commercial purposes. This part of the "Purpose" moderator was measured by a multiple-choice question with seven different options, allowing more than one answer. The question we asked in the survey was "In which of the following examples would you be willing to accept social media platforms to surveil your activity?" (see Appendix 2, Q18 for all seven options).

An overview of how the moderators were measured can be found below:

Table 2. *Overview of Moderators and Measurements*

| Moderator | | Scale | Measurement | Mean | Std.dev |
|---|---|---|---|---|---|
| 1 | Age | Ratio scales<br>(1=18-24, 2=25-34, 3=35-44, 4 =45-54, 5=55-64, 6=65-74, 7=75-84, 8=85 or older) | *How old are you?* | 2.12 | 1.266 |
| 2 | Gender | Male, Female<br>(1=male, 0= female) | *How do you describe yourself?* | 1.64 | .633 |
| 3 | The purpose of surveillance | 1-5<br>(Strongly disagree–strongly agree)<br>Multiple choice questions with more than one option<br>(7 choices) | 1. *I would allow social media platforms to look through my online activity if it is for the betterment of society? (E.g. to avoid crime, child abuse, hate or discrimination).*<br>2. *In which of the following examples would you be willing to accept social media platforms to surveillance your online activity?* | 3.45 | 1.199 |

### 4.0 Methodology

There are several ways to investigate and explore a research question adequately, and the process includes multiple decisions. This section describes our research strategy and the method for the investigation of this thesis. Based on the findings in the literature review in addition to the research question, this thesis will investigate several hypotheses. By answering and testing these hypotheses, we will be able to form a conclusion on our research question:

> "*When firms have to breach privacy to prevent child abuse on social media platforms, would users be more accepting if message screening of their online activity is conducted by an automatic AI detection tool rather than humans?"*

The thesis aims to understand what drives the *choice* of a human or an AI robot in a scenario of surveillance of messages and online activity. It will be measured by two options, a human, or an AI robot in different settings. A regression design with a logit model will be suitable, as this helps us analyse the data when choosing between two choice options. The regression explains what drives the choice and will therefore help us understand the preferences. In addition to the factors that drive the choice of conductor, we want to find which option is the most preferred in the Norwegian population.

### *Hypothesis*

Previous research has created an understanding of people's reaction to the use of private information, which will be the base for the first two hypotheses. These hypotheses are based on the fact that most people do not mind their personal information being looked into. Since people have been sceptical of AI for a long time, we want to know if the use of AI in a personal and intimate setting will make them change their minds. Since AI is based on programming there is a risk that it misinterprets information, which can make people anxious. Therefore, we want to see if this is a factor that affects the choice of conductor. AI may be perceived less judgmental and intimidating when message screening, and it will be interesting to check if this theory holds. Both humans and AI robots can be discriminatory and biased when deciding. However, AI's lack of feelings can be both negative and positive. Therefore, we want to test if the impartiality of a machine can be beneficial and even a better solution when screening personal information. Based on this, we will start by testing the following three hypotheses:

> **H1:** Users will be more willing to let the SoMe platforms assess their messages if the information is processed by AI instead of humans.
>
> **H2:** Users will be anxious of being misinterpreted by AI when screening their personal messages on SoMe.
>
> **H3:** Users of SoMe platforms perceive algorithms as relatively non-discriminatory because they lack emotional experience and feelings, which makes the screening less embarrassing and personal.

The following hypothesis aims to understand which factors can change people's choices. We want to know if different scenarios involving crime prevention on SoMe will change people's choice between AI robots and humans. Here, people's acceptance of surveillance will be tested and linked to the purpose of the surveillance. We want to know if people will be more accepting of surveillance if the purpose is to prevent crime rather than for commercial purposes. In addition, if they are willing to overcome their fear of surveillance, if the purpose is to improve children's safety. We assume that most people want to help society, but that their self-interest may overshadow their morality if their privacy is threatened. The fourth hypothesis is the following:

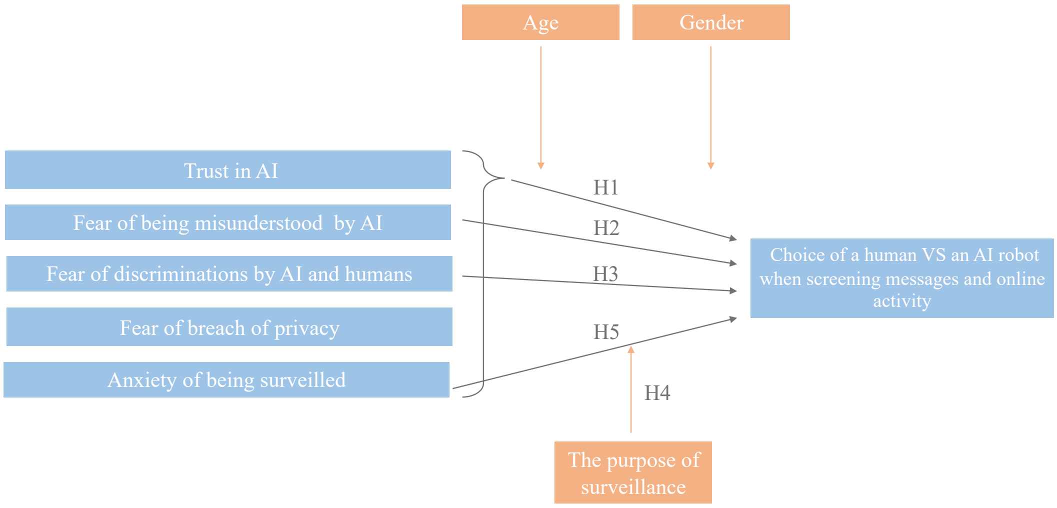
> **H4:** Users will be more willing to allow SoMe platforms to analyse their online activity if the cause is to protect the safety of children rather than a commercial purpose.

The last hypothesis is closely connected to our research question. We want to test if the user's anxiousness and fear of AI technology will affect their choice of the conductor, human or AI. We are investigating if the fear of surveillance makes people prefer humans. Therefore, the connection to be tested is as follows:

> **H5:** Users do not like to have their messages screened by SoMe platforms because they are anxious that the surveillance of online activity can be used against them in a case of child abuse.

To visualise the connection between our DV and the hypotheses we have included the hypotheses in the conceptual model (see Figure 2). It shows the connections that we want to investigate in this thesis.

Figure 2. *The conceptual model with hypotheses*


Age
Gender
Trust in AI
Fear of being misunderstood by AI
Fear of discriminations by AI and humans
Fear of breach of privacy
Anxiety of being surveilled
H1
H2
H3
H5
H4
Choice of a human VS an AI robot when screening messages and online activity
The purpose of surveillance

## 4.1 The survey

The evidence collected in this thesis is based on an extensive survey. To collect the primary data, we used a quantitative method for data collection, as it provided good accuracy and “close-ended” answers. The main aim of the survey is to get a grip on people's preference of conductor of online message screening and people's general acceptance of AI.

### *4.1.1. Design of the survey*

To test our hypotheses and answer our research question, we have used a survey consisting of 21 elaborating questions. We ended up with 108 respondents aged 18 to 74. The aim of the survey was to get insight into how participants react to different questions and scenarios that can help us answer our research question.

*Pre-test*

To get valid results, we did a pre-test on a small, selected group of people to check whether the questions were understandable, if the length of the survey was suitable, and if anything was missing. From the feedback, we added definitions to some of the questions, making it easier for the participants to understand what we were asking. We also removed some questions and changed the measuring scales. The open questions were deleted, to ensure valid and measurable results. After the pre-test, the survey consisted of questions with a mix of the non-cooperative Likert scale 1-5 and 1-7, in addition to multiple-choice and some dichotomous questions.

*The survey*

We divided the survey into three different sections where we chose the order carefully to not reveal too much at the beginning. Thus, the participants thought that the survey was about their online activity and privacy on social media platforms. They were unaware of the actual purpose of the survey, which made them answer more genuine. The survey started with three questions regarding people's online activity on SoMe platforms to better understand their use and relationship to different SoMe platforms. Afterward, we asked about people's privacy and preferences and questions regarding people's anxiety about being judged. The following section handles AI, and people's preferred choices, both with and without a specific scenario.

The following two scenarios was presented with a choice between an AI robot and a human:

> **Scenario 1:** *If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity, like your messages and shared photos?*
>
> **Scenario 2:** *The spread and sharing of illegal materials online are increasing. The European Commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile?*

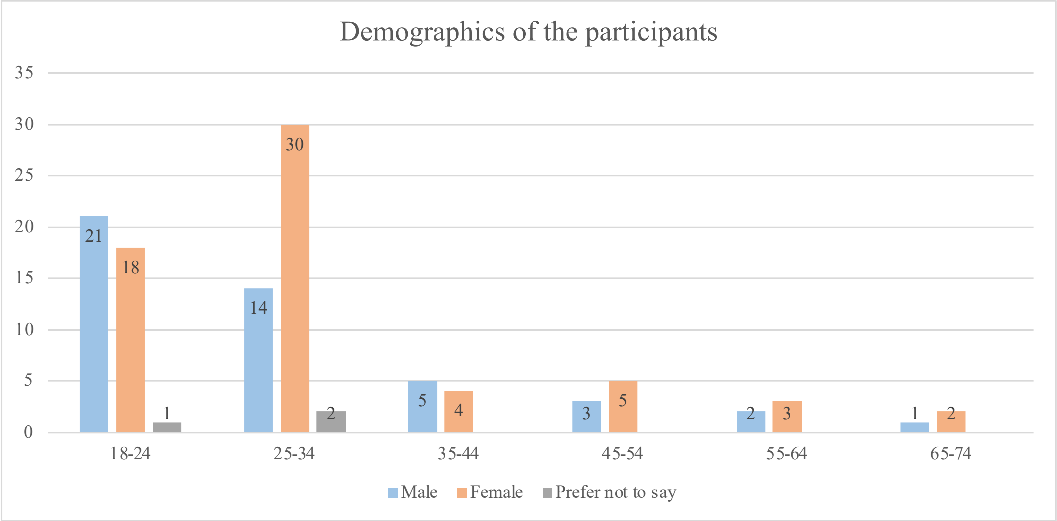
These questions gave us insight into people's general preferences and an overall understanding of people's acceptance of AI in message screening. Scenario 1 was asked first and is a general question about their preferred conductor. Scenario 2 puts the message screening into a scenario of crime prevention. When including both questions, we were able to see if the preferred conductor changed when the purpose of the message screening was for the betterment of children.

At the end of the survey, we asked the participants about their age and gender to be better able to analyse the data. We have provided the participants with three gender options: Male, Female, and Non-binary/Third gender, and a fourth choice to not disclose their gender.

### *Participants*

We want the survey to be neutral and a good representation of the population. The only qualification for taking the survey was above 18 years old. To ensure all the participants were of legal age, we included a tick-off box to confirm that they were above 18 years old. This thesis is restricted to the Norwegian population due to the little research conducted on online crime prevention in Norway. The survey was shared on the author's Facebook and LinkedIn profiles twice to reach out to participants, as the main followers there are Norwegians. The survey was online for three weeks and reached 177 people. However, only 111 people completed the whole survey and passed the attention check. An illustration of our participant's demographics is showed below in Figure 3:

Figure 3. *Demographic of the participants in the study*



### *4.1.2 Data collection*

To use the dataset for analysis, there were several procedures and preparations we had to do. We started by removing all the participants who did not complete the survey and the data from the pre-test. The dataset did not contain any extreme values or outliers. Therefore, we could disregard these as we prepared the dataset. We originally had four different gender options; male, female, non-binary/third gender and prefer to not say. Since no participants chose the "non-binary/third gender", and only 3 participants chose "prefer not to say", we ended up using only males and females. When eliminating irrelevant data, the size of the dataset shrank, and we ended up with a total of 108 participants. The conducted preparations helped us ensure a valid and reliable dataset for further analyses.

### *4.1.3 Validity and reliability*

We selected a survey because it is a quick and anonymous way to understand how people think and behave. We have tried to develop a good survey design to seek both validity and reliability. It is essential to measure validity and reliability to reflect the concept of the tested theory to avoid invalid or biased conclusions. The data must be valid and reliable for the research data to be of value and use (LoBiondo-Wood & Haber, 2014).

### *Validity*

One general challenge with surveys is ensuring validity; therefore, a good survey design is necessary. Validity is defined as "to which an empirical indicator of a concept actually represents the concept of interest" (Goldstein & Simpson, 2015, p. 149). One of the problems with using surveys is content validity, which is ensued by procedures to construct items for a test (Goldstein & Simpson, 2015). The pre-test helped us know *what* and *where* to make changes in order to ensure validity. It was essential for us to create good questions that reflected the research issue and make sure that key-related subjects were not excluded. To ensure a common perception of the survey, we defined possibly unfamiliar terms like *artificial intelligence*. The purpose was to ensure everyone had the same understanding of the terms we used. Different factors may affect the survey results and the degree of truth the answers provide. Since we cannot control how the survey was conducted, external factors that could affect the responses are not considered. However, we tried to ensure internal validity by asking questions to identify factors that could influence the outcome we wanted to research. In that way, we aimed for results that represented the truth in the population.

Since the survey was shared on Facebook and LinkedIn, it is difficult to know if all the participants answered truthfully or took the survey seriously. A selection bias may also occur due to only reaching out to our friends, family, co-students, colleagues, friends, etc. There is a common denominator in all of them: they somehow have a connection to us. We see that the data is skewed towards the younger segment (see Figure 3), which can weaken our results, making them only valid for a specific age group and reducing our external validity.

### *Reliability*

Reliability is described as "the reproducibility of an empirical measure (e.g., internal consistency of the items in a scale, reproducibility of a measurement on different occasions or agreement between raters)" (Goldstein & Simpson, 2015, p. 149). Reliability is important because it defines to which extent the survey will provide us consistent and replicable findings and enable other researchers to make similar observations and conclusions later.

One challenge when asking people about vulnerable topics such as the prevention of OSCEA is that they might answer in a socially acceptable manner instead of what they believe is correct. This is known as the social desirability bias, which can weaken the reliability of the study (Nederhof, 1985). We stressed the importance of honest answers before starting the survey to prevent this bias. The research topics in this thesis were also often unfamiliar to people, hence, their background knowledge of the topics could impact their answers.

Since our survey was quite long, many people did not finish it. The lack of people who finished the survey harmed our results. Another factor to consider is that we did not pay any of the participants. We would probably have gotten more answers if we had offered to pay them; hence more people would have finished the survey. Additionally, if we had been able to share the survey on other platforms with a large follower base, we could have reached more participants. However, the survey did reach people of different ages and genders, seeking reliability.

To check the survey's reliability, we used the Cronbach's Alpha formula on all questions except the multiple-choice questions. Cronbach's Alpha is most commonly used when you have multiple Likert questions in a survey forming a scale, and wish to determine if the scale is reliable. It is viewed as the most appropriate measure of reliability when using questions with Likert scales. (Whitley, 2002) We included all the questions except the 4 that were measured by multiple choice, which gave us the score .858 (see Appendix 3). Since the score is above .70 can we conclude that the survey is of high reliability according to Hinton et al. (2004).

### *4.1.4 Data analysis*

In this study a Bivariate Logistic Regression is used to explain what drives the choice between an AI robot or a human in message screening. Since we only have two possible outcomes (AI or Human), a binary logistic regression is suitable. We call the DV "Choice of conductor", and it will be explained by the IVs. We have chosen to use both Scenario 1 and Scenario 2 as DV in our study. It is used to analyse the likelihood of choosing AI or human given multiple factors that may drive this choice (privacy concerns, age, gender, betterment of the society, purpose of surveillance, etc.). Below, you can find the regression line that we will use with Scenario 1 and Scenario 2 as our DV.

$$Logit\ (Choice\ of\ conductor) =$$

$$\beta_0 + \beta_{Trust_{AI}} x_1 + \beta_{Fear_{Misunderstood}} x_2 + \beta_{Fear_{Discrimination}} x_3 + \beta_{Fear_{Privacy}} x_4 + \beta_{Anxiety_{Surveilled}} x_5 + u$$

## 5.0 Results

This part will present the empirical findings from the research. The analysis will be conducted in SPSS. Qualtrics, where the survey was made, provided a good overview of the collected data, but to conduct a sufficient analysis, we transferred the data into SPSS. Before looking deeper into the dataset, we started by running some descriptive statistics to look at the results.

### ***Differences between the participants***

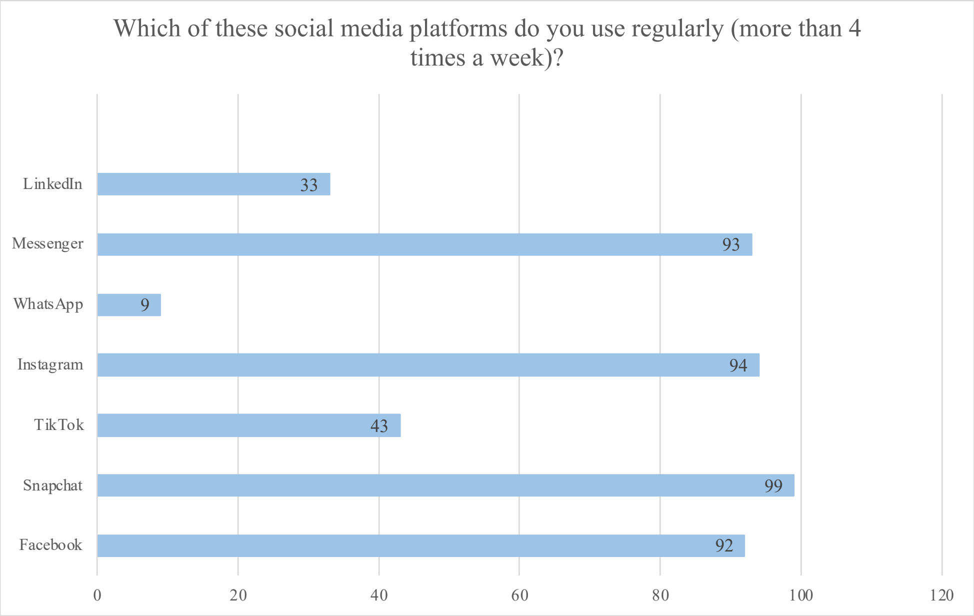
#### *Demographics*

Descriptive statistics with the frequencies function were used to analyse the different age groups and genders. The survey consisted of participants between 18 and 74 years. The majority in the group were under 34 years old (76,9%), while the remaining part were older than 35 years (23,1%). The participants consisted of 62 females (57,4), 46 males (42,6%).

#### *Social media platforms*

We investigated which SoMe platforms are the most used by the participants to understand their ability to relate to the dilemma presented in this thesis. Descriptive statistics showed that the most used SoMe platform is Snapchat followed by Instagram, and Messenger (see Figure 4). A high percentage of the participants using the relevant SoMe platforms make them capable of relating to the presented scenarios of crimes on SoMe platforms. We assumed that all the participants were using Facebook and/or LinkedIn, as these were the only two places the survey was published. The results confirmed this statement, as the majority of the participants use Facebook.

Figure 4. *Illustration of which SoMe platforms the participants use (could choose more than one alternative)*



We also checked the participant's screen time on SoMe in relation to their *gender* and *age* to prove our assumptions and map their use. Since "Gender" and "Age" are two of our moderators, it was important to investigate if it impacted the participant's choice of conductor. The findings show that 3-2 hours on SoMe per day is the most common, but about 7% of the participants use more than 6 hours per day (see Figure 5). Our study also shows that girls use more time on SoMe platforms than boys. The results were as expected and in line with previous research findings. We also checked if "Age" had an impact on time spent on SoMe platforms. From our data, we see that the younger segments use more time on SoMe than the older segments (see Figure 6). These findings support our assumptions, and we look further into the impact on the participant's preferences.

Figure 5. *Time spent on SoMe for the different genders*

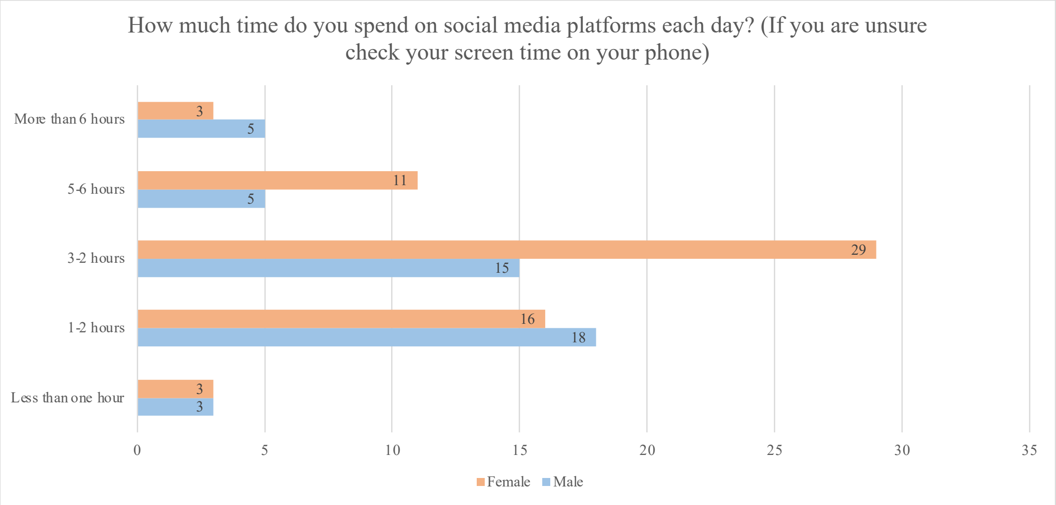
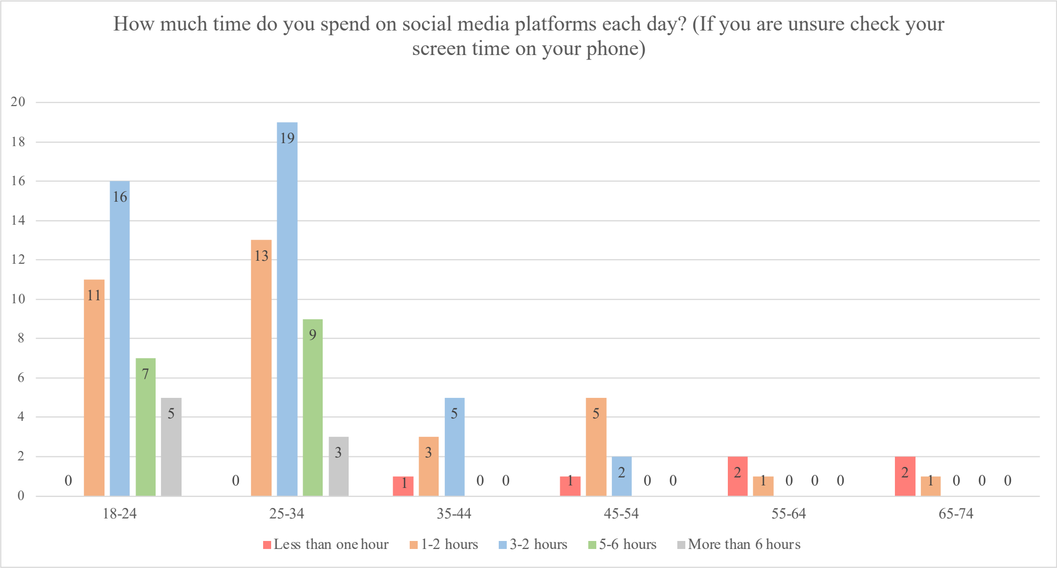


Figure 6. *Time spent on SoMe for the different age groups*



## 5.1 The choice between a human and an AI robot

To better understand the data, we ran a logistic regression explained in the method section 4.1.4, to see if there was a significant difference between the preferred type of conductor (AI or human), and the IVs. Since we were interested in people's choice of conductor, both with and without a scenario, we made two different scenarios presented in section 4.1.1.

The IVs we chose were selected because we believe they can help us explain what drives the choice between AI and human. First, we will start by analysing Scenario 1 and present the results. Then we will do the same to Scenario 2. In the end we will compare the two scenarios and look at how they differ. At the end we will look at other relevant findings from the survey.

### *5.1.1 Scenario 1*

We ran a bivariate logistic regression with Scenario 1 as our DV and coded it as a dummy: (AI=1, Human=0). We wanted to test how the different IVs affected our regression, and if they can explain what drives the choice between AI and human. First, we ran a multiple logistic regression analysis with all our 5 IVs.

Since calculating a $R^2$ is impossible for a logistic regression we must use approximations like Nagelkerke pseudo-$R^2$ and Cox & Snell $R^2$. From the model summary we found a Nagelkerke $R^2$ = .394 and a Cox & Snell $R^2$= .258. The Nagelkerke $R^2$ ranges from 0 to 1 and is often preferred to find the model fit because the pseudo-$R^2$ is typically low. Since the pseudo-$R^2$ gives the predictive power of the model given the assumptions, we can say that the variables have a reasonable model fit. From the Chi-Square test ($\chi^2$=30.751, p <.001), we can draw the conclusion that including these explanatory variables to the model makes sense. The classification table shows the logistic regression models classificatory power and reports the overall percentage of correctly classified cases. We see that the overall correct percentage is 84.5% (see Appendix 4).

From Table 3 we can see that only IV 3 have a significant effect on the preferred choice of conductor (p < .05). On the other hand, IV 1 and IV 4 are not too far off, but they still do not have a significant impact on the choice (p > .05) when using a 5% significance level. However, they are significant if we use a 10% significance level. IV 2 and IV 5 have the least significant effect on the model (see Table 3). There is a possibility that multicollinearity-problems makes the IVs less significant. The beta coefficients in the model gives us the opportunity to define the direction of a change of the logarithm of the odds. From Table 3 we can see that IV 1 and IV 2 have a negative effect on our DV, while the other IVs have a positive effect, when looking at the beta coefficient. The negative effect indicates a decrease on the DV for a unit change in the IV.

The result from the model is described in Table 3 below:

Table 3. *Multiple logistic regression with all 5 IVs on Scenario 1*

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 1 | Trust in AI | -.700 | .406 | 2.966 | 1 | .085 | .497 |
| 2 | Fear of being misunderstood by AI | -.144 | .253 | .325 | 1 | .568 | .866 |
| 3 | Fear of discrimination by AI and humans | 4.037 | 1.159 | 12.119 | 1 | <.001 | 56.517 |
| 4 | Fear of breach of privacy | .434 | .243 | 3.181 | 1 | .075 | 1.543 |
| 5 | Anxiety of being surveilled | .374 | .270 | 1.920 | 1 | .166 | 1.453 |
| | *Constant* | -5.793 | 2.240 | 6.634 | 1 | .010 | .003 |

We found that not all the IVs had a significant effect on the model, when included in the multiple regression model (see Table 3). Therefore, we ran several simple regressions on each variable alone, to see what happened. We see that *IV 2*, *IV 4* and *IV 5* still do not have a significant impact on the choice (p > 0.05) (see Appendix 5, 6 and 7). The Nagelkerke $R^2$ was low for all the variables and we can say that these variables are not contributing to explaining the choice of conductor. Further, the results for IV 1 and IV 3 was improved by a linear regression.

*IV 1: Do you trust AI (alone)*

We started by testing IV 1 to see the model fit. The findings show that IV 1 by itself has a significant effect on the model (p= .033 < .05). From the model summary we found a Nagelkerke $R^2$ = .044 and we can say that this variable has a reasonable model fit on its own (see Table 4). We expected this variable to work better alone due to problems with multicollinearity.

Table 4. L*inear logistic regression with IV 1*

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 1 | Trust in AI | -.665 | .312 | 4.532 | 1 | .033 | .514 |
| | *Constant* | .816 | .937 | .758 | 1 | .384 | 2.262 |

*IV 3: Fear of discrimination by AI and humans*

This IV by itself has a significant effect on the model (p= .001 < .05). From the model summary we found a Nagelkerke $R^2$ = .330 which is the highest of all the IVs. Hence, we can say that this variable explains the model well. It indicates that IV 3 is the IV that helps explain people's choice between AI and humans (see Table 5).

Table 5. *Linear logistic regression with IV 3*

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 3 | Fear of discrimination by AI and humans | 3.977 | 1.086 | 13.410 | 1 | <.001 | 53.333 |
| | *Constant* | -5.651 | 1.190 | 22.538 | 1 | <.001 | 0.004 |

*Including moderators*

Further, the moderators "Age" and "Gender" that apply for all the hypotheses H1-H5 will be included in the model to see if they can help explain what drives the choice of conductor. We ran a multiple logistic regression with the moderators and ended up with a Nagelkerke $R^2$ = .282 and a Cox & Snell $R^2$ = .203. Both values of Nagelkerke and Cox & Snell decreased when the two moderators were included. This means that the model fit is better without our moderators. The significance level of alle the IVs  are quite similar but have slightly decreased as well. From the classification table we see that the overall correct percentage has increased to 86.4% (see Appendix 8). None of the moderators were highly significant, and we conclude that the moderatos together do not have a significant effect on explaining the choice (see Table 6).

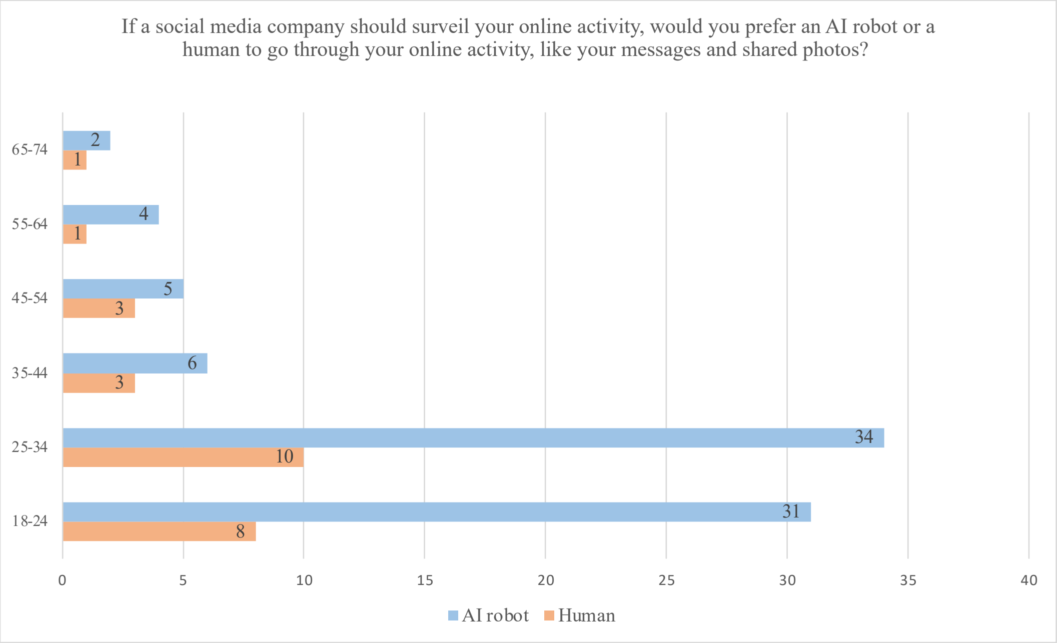
Table 6. *Multiple logistic regression with all 5 IVs and moderator 1 and 2 on Scenario 1*

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 1 | Trust in AI | -.772 | .419 | 3.389 | 1 | .066 | .462 |
| 2 | Fear of being misunderstood by AI | -.171 | .256 | .447 | 1 | .504 | .843 |
| 3 | Fear of discrimination by AI and humans | 4.057 | 1.157 | 12.421 | 1 | <.001 | 57.815 |
| 4 | Fear of breach of privacy | .472 | .246 | 3.683 | 1 | .055 | 1.603 |
| 5 | Anxiety of being surveilled | .376 | .277 | 1.838 | 1 | .175 | 1.456 |
| Moderators | | | | | | | |
| 1 | Age | .142 | .246 | .3.683 | 1 | .055 | 1.603 |
| 2 | Gender | -.422 | .544 | .602 | 1 | .438 | .656 |
| | *Constant* | -5.270 | 2.470 | 4.554 | 1 | .033 | .005 |

### *Moderator effects*

We wanted to test the moderators “Age” and “Gender” further, to see more specific *age* and *gender* differences. We ran an analysis between the age groups and the choice of AI versus humans. The most outstanding results were as follows, in group 2 (25-34 years), 34 people chose AI, and only 10 chose a human (see Figure 7). This is the group with the widest difference out of all the age groups. However, we see a common trend of AI being the most preferred type in all the segments. Even the participants in group 6 (65-74 year) preferred AI, but this trend is not providing valid results alone due to the group only including 3 participants. Thus, the finding show that there is no significant moderator effect of gender since all the age groups chose the same conductor. We conclude that AI is the most preferred choice, but the age and gender do not determine this choice.

Figure 7. *Preference of conductor in Scenario 1*



The moderator "The purpose of surveillance" was tested to see if it can help explain H5 as shown in our conceptual model in section 4.0. This moderator was coded with seven different dummies representing all seven different options in the survey (see Appendix 2). We ran a multiple logistic regression on IV 5 with the moderator, and found that none of the values are significant ($p > .05$). We conclude that the purpose of the surveillance does not help explain why people are anxious of being surveilled (see Appendix 9).

### *5.1.2 Scenario 2*

To investigate the differences between the two scenarios we ran a bivariate logistic regression with Scenario 2 as our DV, similar to Scenario 1. We coded it the same way, as a dummy: (AI=1, Human=0) to see if the IVs can help explain what drives the choice between AI and human.

A multiple logistic regression analysis with all our 5 IVs was conducted first. From the model summary we found a Nagelkerke pseudo-$R^2$ = .256 and a Cox & Snell $R^2$ = .185. From the pseudo-$R^2$ we see that the predictive power of the model given the assumptions, have a reasonable model fit. Chi-Square test gave ($\chi^2$==21.275, p <.001), and we can conclude that including these explanatory variables to the model makes sense. The classification table for this regression shows an overall correct score of 76%, which is lower than in Scenario 1 (see Appendix 10). This is a decrease of 10% when comparing it with Scenario 1.

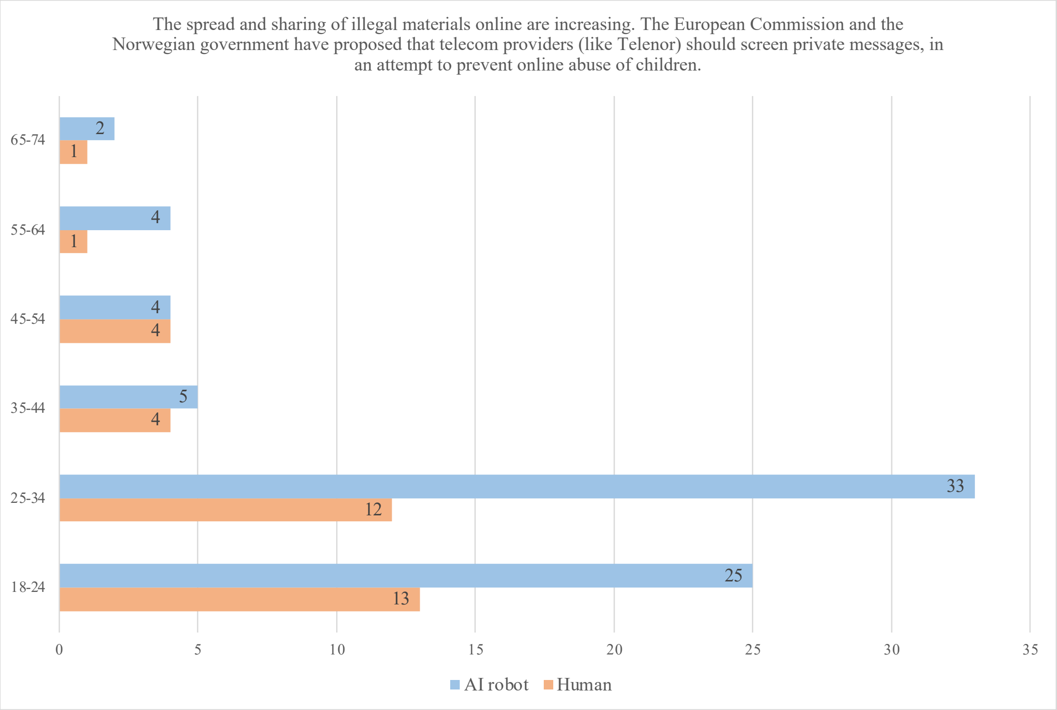
Like Scenario 1, we see that only IV 3 have a significant effect on the preferred choice of conductor (p< .05). None of the other IVs have a significant impact on the choice (p> .05), indicating that using Scenario 2 as our DV makes a less suitable model than using Scenario 1 (see Table 7). We see that IV 1 and IV 5 have a negative effect (negative beta coefficient) on our DV, while the other IVs have a positive effect.

Table 7. *Multiple logistic regression with all 5 IVs on Scenario 2*

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 1 | Trust in AI | -.474 | .327 | 2.097 | 1 | .148 | .623 |
| 2 | Fear of being misunderstood by AI | .030 | .197 | .022 | 1 | .881 | 1.030 |
| 3 | Fear of discrimination by AI and humans | 3.409 | 1.108 | 9.465 | 1 | .002 | 30.238 |
| 4 | Fear of breach of privacy | .239 | .196 | 1.486 | 1 | .223 | 1.270 |
| 5 | Anxiety of being surveilled | -.105 | .211 | .246 | 1 | .620 | .901 |
| | *Constant* | -3.592 | 1.844 | 3.794 | 1 | .051 | .028 |

Since the moderators showed a minimal effect on the model in Scenario 1, we have decided to not include them in this model (see Appendix 11). The classification model only gave an overall correct score of 74% (see Appendix 12). The results show that there might be a problem with multi collinearity, providing problems in the model. However, we tested how the moderator "Age" affected Scenario 2, to enable a comparison between the scenarios. From Figure 8 we see that the most outstanding results were in group 4 (45-54 years) where the genders preference between AI and humans were equal. In all the other age groups there are no gender differences, as AI is dominating as the most preferred choice. As in Scenario 1, the finding show that there is so significant moderator effect of gender.

Figure 8. *Preference of conductor in Scenario 2*



### *5.1.3 Comparing Scenario 1 and Scenario 2*

After presenting both scenarios with different combinations in the regression to find the best model we will compare them. We see that using Scenario 1 as our DV provided the best model fit. There seems to be a problem with correlation between some of our IVs which makes gives the weak significance scores. However, when we tested the IVs separately, we see that the significance is improved, especially for IV 1 and IV 3. The reason may be that the three IVs regarding fear are too similar.

In the figures showing the moderator effects, we can see how the different ages have responded in both scenarios (see Figure 7 and Figure 8). One interesting common observation is that in all age groups, is that AI is the most preferred choice for both scenarios. In Scenario 2 we found that most people in the youngest segment converted from AI to human. The finding indicate that the youngest segment is the most afraid of AI when the stakes increase (Figure 8). However, gender do not have a significant impact on the preferred conductor in neither of the scenarios as a moderator. We will now look at the different age groups preferences of conductor.

Descriptive statistics was used to understand the preference of conductor. The figures below show how the different genders responded between AI or human in the two scenarios (see Figure 9 and Figure 10). In both scenarios, men have the highest score for choosing AI. 78% of the males choose an AI robot as the preferred conductor in Scenario 1, and 74% in Scenario 2. On the other hand, females seem more sceptical towards AI robots. In Scenario 1, 74% of the females chose AI, and 63% in Scenario 2. When Scenario 2 was presented men's preference of AI dropped with 4%, while female's preference dropped with 11%. We see a gender difference that may indicate that females are more sensitive than men in morally relevant situations.

Figure 9. *The preferred conductor based on gender on Scenario 1*

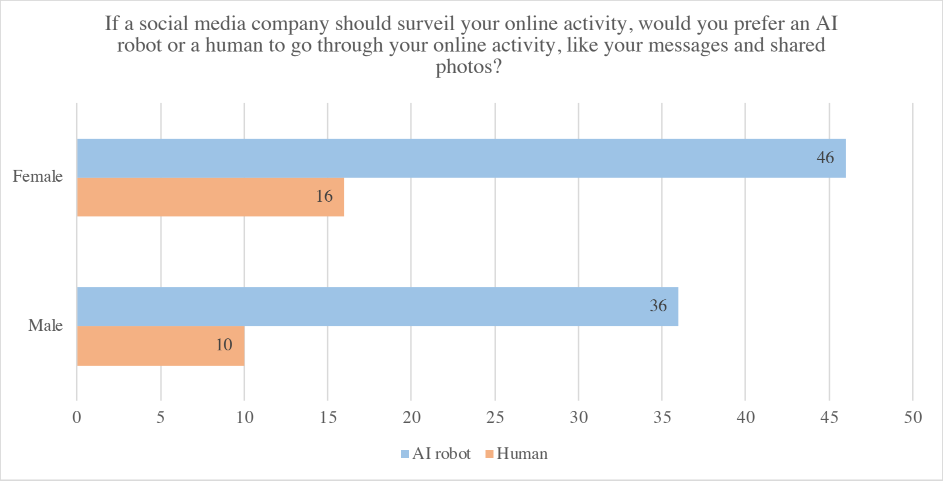
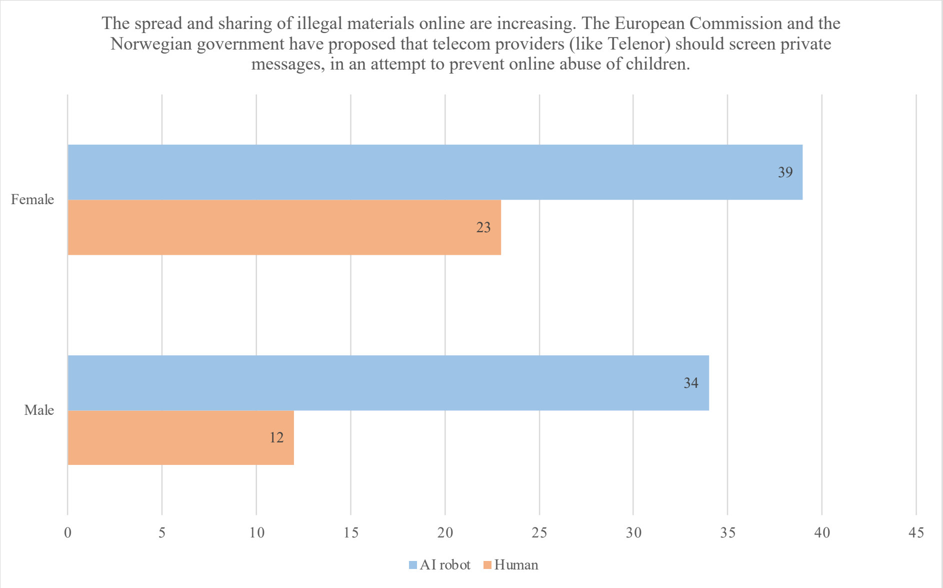


Figure 10. *The preferred conductor based on gender on Scenario 2*



We will now look at some of the other general findings from the survey, to understand what drive the choice of conductor.

*Other findings from the survey*

Since both scenarios indicated weak moderator effects, we wanted to further check if the collected answers from our survey that issued the purpose of the surveillance could help explain how the purpose may affect the choice of conductor. From Figure 11 below we see that approximately 60% somewhat or strongly agreed to the assumption. This result shows that the acceptance of surveillance is high when the purpose of the surveillance for the betterment of society. However, about 1% strongly disagree to surveillance even if the purpose has good intentions. To elaborate this further, we created Figure 12 which shows in what situations they would accept surveillance. The results indicate a significant difference with a higher preference for crime preventive purposes rather than commercial purposes.

Figure 11. *The purpose of surveillance*

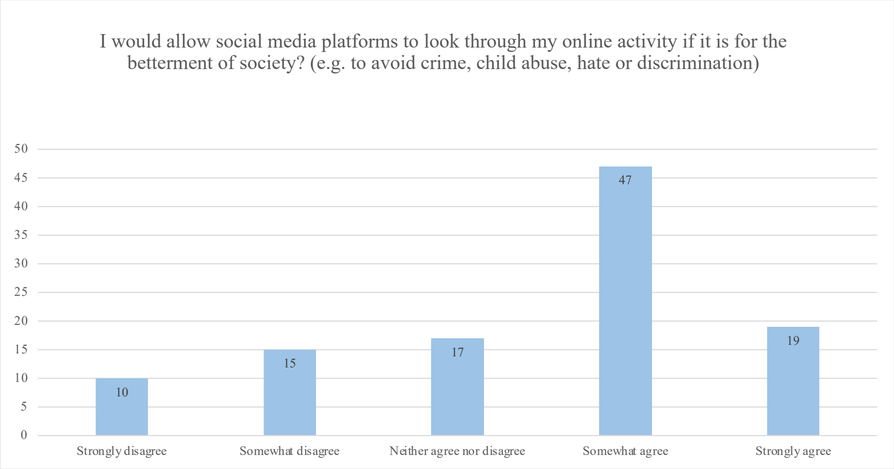
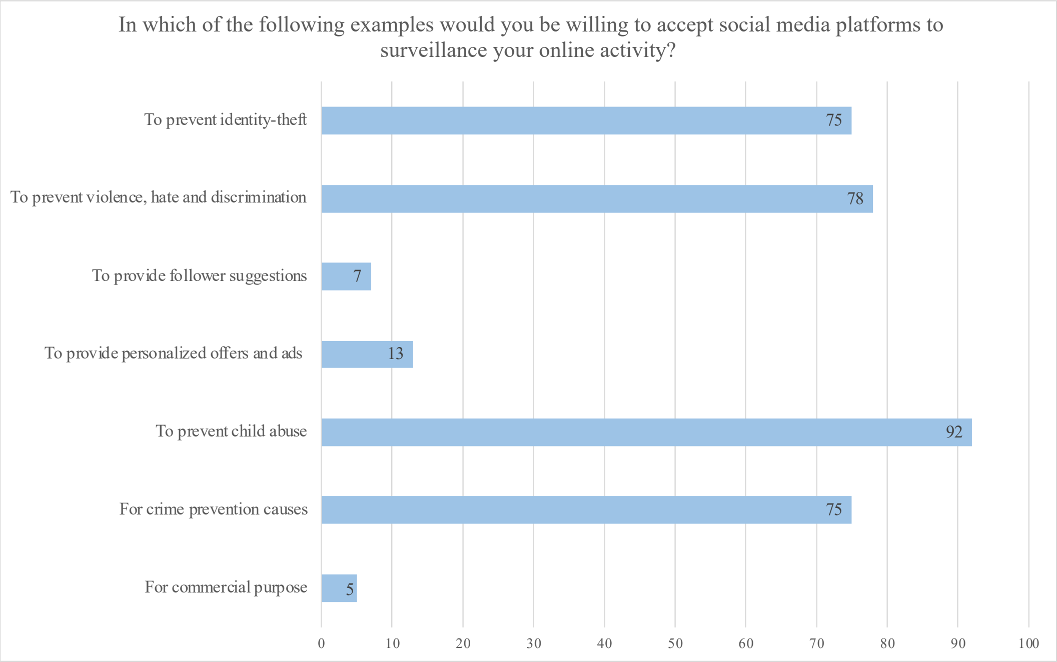
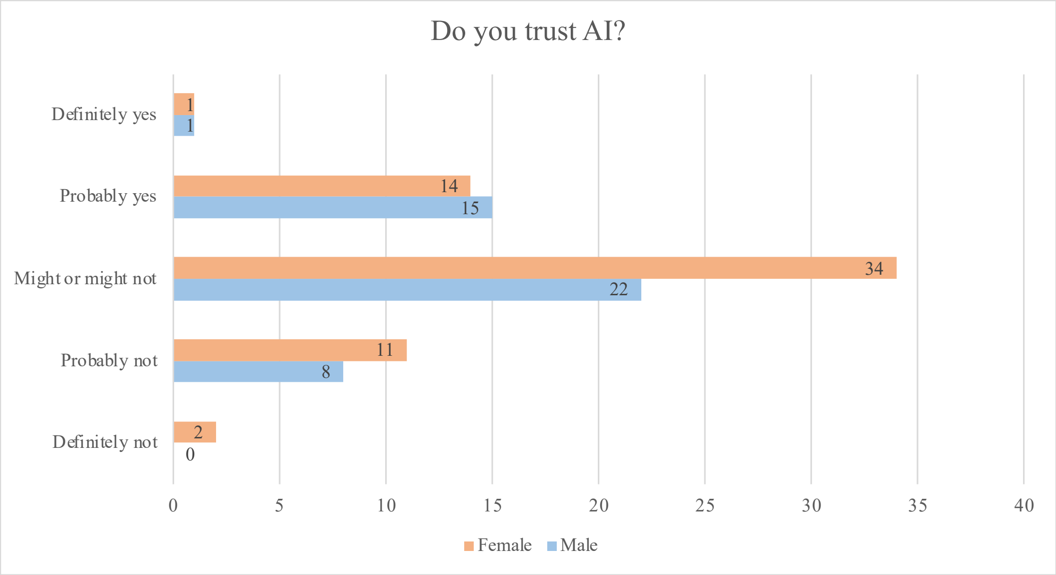


Figure 12. *Situations where surveillance is accepted by the participants*



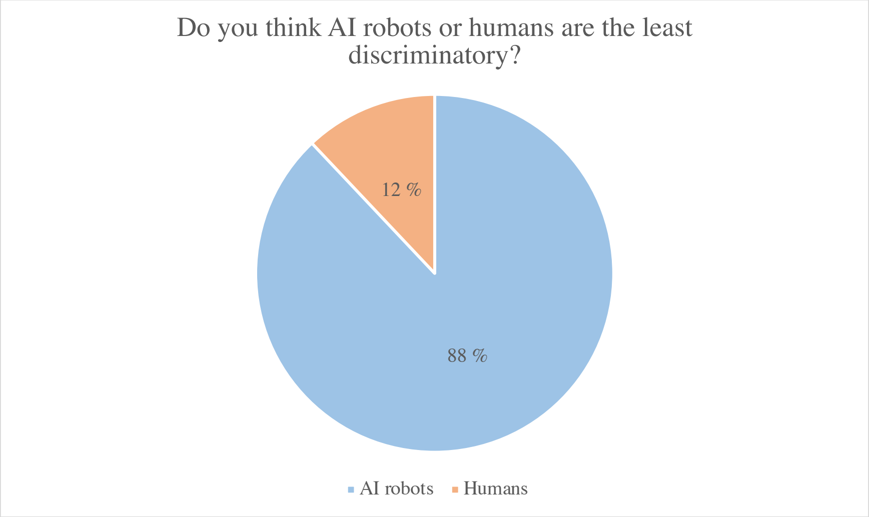
We find it relevant to look at peoples trust in AI, which is coded (1 = definitely not – 5 = definitely yes) as explained in section 3.1. As Figure 13 below shows, over 50% answered "Might or might not," which gives us nothing. However, the larger part of the remaining 50% chose "probably yes or "definitely yes", indicating that they trust AI. Most people trusting AI support our previous findings of AI as the preferred conductor of message screening.

Figure 13. *Trust in AI*



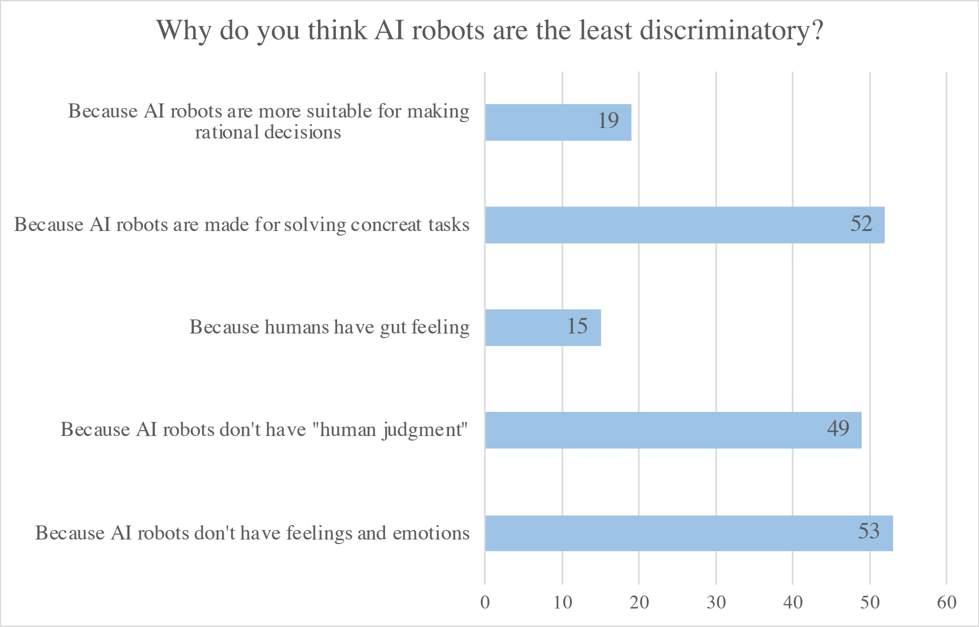
The next question that we looked into was regarding whom is perceived as the least discriminatory out of AI and humans. As Figure 14 below shows, the result was clear on AI being perceived as the least discriminatory by 88% of the participants. This finding also supports the assumption of AI to be the most preferred conductor.

Figure 14. *Which conductor is the least discriminatory*



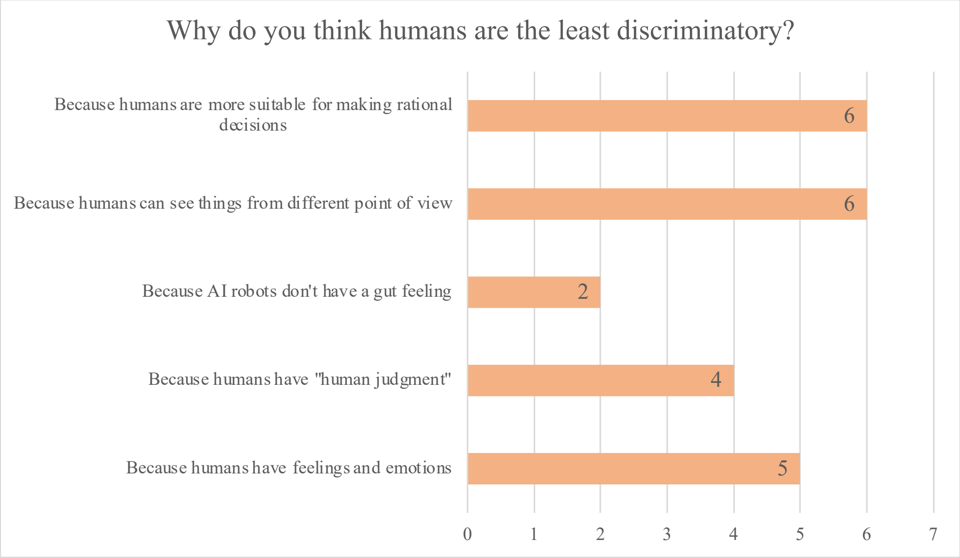
To further understand why AI is perceived as the least discriminating conductor, we look at the responds on the question “Why do you think AI robots is the least discriminatory?” from the survey. It was possible to choose several of the assumptions in this question. This multiple-choice question was only given to the 88% of the participants chose AI robots to be least discriminatory. 53 of the participants mean that AI robots are the least discriminatory because they do not have feelings and emotions, which supports our third hypothesis (see Figure 15).

Figure 15. *Reasons why the participants find AI the least discriminatory*



We also looked deeper into the reasons why some people chose humans to be the least discriminatory. One surprising finding was that the majority believed it was because of their existing feelings and emotions (see Figure 16). Interestingly, the driver of both choices AI and humans, are completely opposite. AI is preferred because they do not have feelings, and humans are preferred because they have feelings.

Figure 16. *Reasons why the participants find humans the least discriminatory*



## 6.0 Discussion

We will now look at how our findings can help us answer our hypothesis presented in section 4.0.

Using two scenarios enabled us to test our hypotheses in a sufficient way. We will now present if the hypotheses can be accepted or not.

H1: *Users will be more willing to let the SoMe platforms assess their messages if the information is processed by AI instead of humans.* This hypothesis was the only one we expected to be explained by all the IVs. As our regression showed, the model has a medium good fit in both cases, but some issues presented above occurred. However, the preference of AI above humans was clear and IV 3 was highly significant in all cases. Since the majority of the participants chose AI, our findings support this hypothesis, and we conclude that the statement is correct.

*H2: Users will be anxious of being misinterpreted by AI when screening their personal messages on SoMe.* This hypothesis var tested by IV 2 on peoples fear of being misunderstood by AI. As the results showed, this IV had no significant impact on the model. We conclude that we cannot accept this hypothesis.

*H3: Users of SoMe platforms perceive algorithms as relatively non-discriminatory because they lack emotional experience and feelings, which makes the screening less embarrassing and personal.* This hypothesis was based on IV 3, which was one of the most significant variables. It shows the fear of being discriminated against affect people's choice of conductor. Therefore, we conclude that this hypothesis is supported in our thesis.

*H4: Users will be more willing to allow SoMe platforms to analyse their online activity if the cause is to protect the safety of children rather than a commercial purpose.* The moderator "The purpose of surveillance" was the driver of this hypothesis. As multiple of our findings show, the purpose of the surveillance had a huge impact on people's acceptance of surveillance. Thus, we conclude that this hypothesis is supported.

*H5: Users do not like to have their messages screened by SoMe platforms because they are anxious that the surveillance of online activity can be used against them in a case of child abuse.* IV 5 was investigating this hypothesis. As the results showed, there was not a significant effect from this IV on the model. However, the purpose of the surveillance did matter in the level of acceptance towards surveillance. Still, we must conclude not to support this hypothesis due to the lack of proof regarding anxiety of surveillance.

Table 8. *Summary of Results*

| Hypotheses | Variables | Results |
|---|---|---|
| H1 | 1. Trust in AI<br>2. Fear of being misunderstood by AI<br>3. Fear of discrimination by AI and humans<br>4. Fear of breach of privacy<br>5. Anxiety of being surveilled | *Supported* |
| H2 | 2. Fear of being misunderstood by AI | *Not supported* |
| H3 | 3. Fear of discrimination by AI and humans | *Supported* |
| H4 | 5. Anxiety of being surveilled with the moderator "The purpose of surveillance" | *Supported* |
| H5 | 5. Anxiety of being surveilled | *Not supported* |

### *General discussion*

As the results show, there are many factors that contribute to the choice between an AI robot or a human in a potential crime prevention situation. We have investigated which conductor is most preferred. Our thesis showed that AI is the most preferred in all age groups and for all genders. The results are as expected based on the literature review. We know that some people will be sceptical of AI and that the scepticism will increase along with the stakes of the task. Both scenarios also confirm hypothesis H1, of a general preference for AI above a human in SoMe message screening if the situation involves high stakes.

However, we have only tested it for two scenarios related to the topic of interest in this thesis. One interesting finding was the importance of the purpose of the message screening. If it was to prevent crime, discrimination and hate comments, the majority had more acceptance of the surveillance. The reason can either be that people feel morally obligated to answer in a sufficient matter, because they know that it is expected from them. If the question was asked in reality and not trough a survey, the answers may have changed. In addition, if the scenarios had been different, it could have affected the acceptance of the screening. Another view is the perception of oneself. If you know that you have something to hide, you will be less likely to accept surveillance of any kind, comparted to a person that has never committed a crime.

Our findings also show that as the stakes of the decision increases (by introducing two scenarios), the participants tend to have a larger preference for a human. This is in line with previous research on how humans prefer humans to make morally relevant decisions and especially in situation where they are afraid to be misinterpreted. As the results showed when we ran the regression on the relationship between which is the least discriminatory of AI versus humans, the answer was AI. This finding explains why most people want AI to screen their messages. The logical explanation will be to believe that the fear of discrimination plays a big role in the choice of conductor, pointing towards a preference for AI. As we elaborated in the beginning of the thesis, AI can be a non-judgmental third party without any intention of harming or humiliating the users. It is logical to expect AI to be the preferred choice, if it is perceived as less discriminatory and judgmental. However, we saw that some people are and probably always will be sceptical of new technologies.

## 7.0 Conclusion

The aim of the thesis was to investigate what drives the choice between an AI robot or a human in a setting of message screening for prevention of OCSEA.

The inclusion of AI into decision making can be challenging in handling biases and making the users trust AI. SoMe companies must be aware of the potential resistance and work towards a common understanding that can benefit society and the users' safety online. The attention of different Governments and media awareness regarding this issue can help make the increasing problem of OCSEA understood and stopped. Another essential problem is the conflict between using users' data for crime prevention and how this affects the users' privacy and the companies' reputation.

The results of the survey showed a clear preference in AI robots above human in all ages and genders. When looking deeper into the drivers of the choice, we found different drivers that had an effect on why people prefer AI above humans. The driver that had the largest impact was the participant's impression of which conductor was the least discriminatory. Most people believe that AI will be less judgmental and discriminatory, and that is the main reason why people would prefer AI above humans in a personal and possibly embarrassing screening situation. However, some people are sceptical of new technologies like AI. To overcome this obstacle, it will remain important that governments all over the world contribute to make the prevention of OCSEA a priority. Even though not all of the used IVs were significant, we conclude that there is enough evidence to say that AI is the preferred conductor of message screening in a situation of crime prevention answering our research question.

## 8.0 Theoretical implications and further research

All studies have challenges and limitations. The biggest challenge in this thesis was the use of different scales and answer-types in the survey. It made it challenging to compare the answers and to use all the questions we wanted in different setting. It limited us in the way we wanted to explore the data. However, the required adjustments made the data possible to handle eventually, by coding the dataset properly.

The research topics in this thesis were also often unfamiliar to people, and their background knowledge of the topics could impact their answers. Therefore, one limitation in the study was that we did not know the participants background knowledge and technological skills. AI technology is something that can be unfamiliar to people, which can make the answers not valid if they do not understand the meaning of the question or the technology. In addition, a wider spread of age in the used population could have made the answers different. This study was skewed towards a young segment. Therefore, further research should include a larger and less skewed group of people.

Another challenge was that the survey was quite long, with an average take time of 10 minutes. One problem that occurred was getting all the participants to complete the survey. It made it challenging to get enough answers, as some people exited the survey before completing it. We should have thought of this and that could easily be fixed by making all the questions mandatory or making the survey shorter. However, on the first page of the survey, we warned the participants about the take-time. Surveys can also present a challenge when asking people about vulnerable topics such as the prevention of OSCEA. The participants might answer in a socially acceptable manner instead of what they believe is correct, which can make the answers invalid. It would have been interesting to try a different method for data collection, for example an experiment to see how the results differ, and explain the choice of conductor more sufficiently.

Since the reporting and removing of crime-material on SoMe has been voluntary it has made it hard to stop OCSEA. This study has not contacted the SoMe companies in the investigating, hence cannot know how they are looking at the situation. The new degenerations and laws can affect the SoMe companies' decisions and destroy the platform's reputation, and therefore it would be interesting to get the SoMe companies point of view in further research.

## 9.0 References

Aanerød, L. T., & Mossige, S. (2018). *Nettovergrep mot barn i Norge 2015–2017.* Norsk institutt for forskning om oppvekst, velferd og aldring.

Aaskervik, A.-L. (2020, November 2). *AiBA - Avslører cybergrooming.* Retrieved May 27, 2022, from NTNU Discovery: https://ntnudiscovery.no/aiba-avslorer-cybergrooming/

Abubakar, A. M., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizationalperformance. *Journal of Innovation & Knowledge*(4).

Aguilar, J. F. (2021). *Regulation (EU) 2021/... of the European Parliament and of the Council of ..* Retrieved from May 27, 2022, https://www.europarl.europa.eu/doceo/document/A-9-2020-0258-AM-039-039_EN.pdf

Alshura, M. S., Abughazaleh, M., & Zabadi, A. M. (2018, November). Big Data in Marketing Arena. Big Opportunity, Big Challenge, and Research Trends: An Integrated View. *Management and Economics Review, 3*(1).

Bambauer, D. E., & Risch, M. (2021, November 10). *When do consumers prefer algorithmic versus human decisionmakers?* Retrieved May 3, 2022, from Brookings: https://www.brookings.edu/techstream/when-do-consumers-prefer-algorithmic-versus-human-decisionmakers/

Barth, S., & De Jong, M. D. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and informatics*.

Bateman, T. (2022, January 10). *EU plans to fight child sexual abuse online with new law obliging tech firms to report offences*. Retrieved May 10, 2022, from Euronews.next: https://www.euronews.com/next/2022/01/10/eu-plans-to-fight-child-sexual-abuse-online-with-new-law

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society, 11*(1), pp. 1-5.

Bergen, E. (2014). Comparing adult-youth and adult-adult online sexual solicitation: Manipulative behaviors, situational factors, and outcomes.

Bertrand , M., & Mullainathan, S. (2003, July). *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*. Retrieved May 5, 2022, from National Bureau of Ecnomic Research (NBER: https://www.nber.org/papers/w9873

Bigman, Y. E., & Gray, K. (2018). People Are Averse to Machines Making Moral Decisions. *Cognition*, pp. 3-30.

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2022). Algorithmic Discrimination Causes Less Moral Outrage than Human Discrimination. *Journal of Experimental Psychology: General*, pp. 2-5.

Bucklin, R. E., Lehmann, D. R., & Little, J. D. (1998). From Decision Support to Decision Automation: A 2020 Vision. *Marketing Letters, 9*(3), pp. 235–246.

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2019, October 23). A systematic review of algorithm aversion in augmented decision making. *Journal of behavioral Decision Making*.

Cambridge Dictionary. (2020). *Algorithm*. Retrieved May 5, 2022, from *Cambridge Dictionary*: https://dictionary.cambridge.org/dictionary/english/algorithm

Cambridge Dicitonary. (n.d.). *Social media*. Retrieved May 5, 2022, from Cambridge Dicitonary: https://dictionary.cambridge.org/dictionary/english/social-media

Carrascal, J., Riederer, C., V, E., Cherubini, M., & de Oliveria, R. (2013, May). Your browsing behavior for a Big Mac: Economics of Personal Information Online. *Proceedings of the 22nd international conference on World Wide Web* .

Carroll, A. B. (1999, September). Corporate Social Responsibility: Evolution of a Definitional Construct. *Business & society*, pp. 268-295.

Carroll, A. B., & Schwartz, M. S. (2011, October 11). Corporate Social Responsibility: A Three-Domain Approach. *Philosophy Documentation Center*, pp. 503-530.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019, July 5). Task-Dependent Algorithm Aversion. *Journal of Marketing Research, 56*(5), pp. 809-825.

Ceci, L. (2021, December 16). *Average daily time spent by users worldwide on social media apps from October 2020 to March 2021, by gender.* Retrieved June 1, 2022, from Statista: https://www.statista.com/statistics/1272876/worldwide-social-apps-time-spent-daily-gender/

Christensen, S. (2021, April 26). *Norge i verdenstoppen i bruk av sosiale medier.* Retrieved May 5, 2022, from Kundeserviceavisen: https://kundeserviceavisen.no/2021/04/26/norge-pa-verdenstoppen-i-bruk-av-sosiale-medier/

Citeron, C. L. (2011). The role of information in strategic decision-making. *International journal of information management, 31*(6).

Copeland, B. J. (2022, Mars 18). *Artificial intelligence.* Retrieved May 3, 2022, from *Britannica.*: https://www.britannica.com/technology/artificial-intelligence

Cole, D. (2020). The Chinese Room Argument. *Stanford Encyclopedia of Philosofy*.

Colson, E. (2019, July 8). *What AI-Driven Decision Making Looks Like*. Retrieved June 20, 2022, from Harvard Business Review: https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Council of Europe. (n.d.). *End Online Child Sexual Exploitation and Abuse @ Europe (EndOCSEA@Europe)*. Retrieved June 10, 2022, from Council of Europe: https://www.coe.int/en/web/children/endocsea-europe

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization science, 10*(1).

De Cremer, D., Van Dick, R., Tenbrunsel, A., Pillutla, M., & Murnighan, J. (2011). Understanding ethical behavior and decision making in management: A behavioural business ethics approach. *British Journal of Management*(22).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014, November). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology General*.

Dollarhide, M. (2021, August 31). *Social Media*. Retrieved May 6, 2022, from Investopedia: https://www.investopedia.com/terms/s/social-media.asp

Dorotic, M. (2021, September 7). *Preventing Online Abuse Without Upsetting Customers: Mission Impossible?* Retrieved January 9, 2022, from BI Norwegian Business School: https://www.bi.edu/research/business-review/articles/2021/09/preventing-online-abuse-without-upsetting-customers-mission-impossible/

ECPAT Norway. (2021). *Seksuell utnyttelse og overgrep av barn på internett: Gjennomgang av norsk rettspraksis.* ECPAT Norway.

Ertzeid, H. (2021, May 18). *Slik blir barn og unge utsatt for seksuelle overgrep på nett*. Retrieved May 24, 2022, from OsloMet:

https://www.oslomet.no/forskning/forskningsnyheter/barn-unge-seksuelle-overgrep-nett

EUR-Lex. (2016, April). *Regulations: Regulation(EU) 2016/679 of the European Parlament and of the Council of 27 April 2016.* Retrieved May 24, 2022, from EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

European Commission. (2022, May 11). *Fighting child sexual abuse: Commission proposes new rules to protect children*. Retrieved June 1, 2022, from European Commission: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2976

European Commission. (n.d.). *What is personal data?.*Retrieved May 2022, from European Commission: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en.

Floridi, L. (2016, January 25). *Humans have nothing to fear from intelligent machines*. Retrieved April 10, 2022, from Financial Times: https://philarchive.org/archive/FLOHHN

FN-sambandet. (2022, February 4). *Barnekonvensjonen*. Retrieved June 12, 2022, from FN-sambandet: https://www.fn.no/om-fn/avtaler/menneskerettigheter/barnekonvensjonen

Frankenfield, J. (2021). *Herbert A. Simon*. Retrieved January 6, 2022, from Investopedia: https://www.investopedia.com/terms/h/herbert-a-simon.asp

French, R. M. (2000). The Turning Test: the first 50 years. *Trends in cognitive sciences, 4*(3).

Frøyland, L. R., Solstad, G. M., Andersen, P. L., Tveito, S. B., Folstad, S. H., & Skilbrei, M.-L. (2021). Seksuelle overgrep mot barn og unge via digitale medier. *Velferdsforskningsinstituttet NOVA*.

Ganster, D. C. (2005). Executive job demands: Suggestions from a stress and decision-making perspective. *Academy of Management Review*, 30(3), 492-502.

Gilbert, R; Widom, C S.; Browne, K; Fergusson, D; Webb, E; Janson, S;. (2009). Burden and consequences of child maltreatment in high-incom contries. *The Lancet*.

Goldstein , J. M., & Simpson, J. C. (2015). *Validity: Definitions and Applications to Psychiatric Research.*

Grygiel, J., & Brown, N. (2019, May). Are social media companies motivated to be good corporate citizens? Examination of the connection between corporate social responsibility and social media safety. *Telecommunications Policy, 4*(43).

Gugerty, L. (2006). Newell and Simon's logic theorist: Historical background and impact on cognitive modeling. *roceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*.

Gurung, A., & Raja, M. (2016). Online privacy and security concerns of consumers. *Information & Computer Security*.

Hambrick, D. C., Finkelstein, S., & Mooney, A. C. (2005). Executive job demands: New insights for explaining strategic decisions and leaderbehaviors. *Academy of management review*, *30*(3), 472-491.

Harappa. (2021, June 21). *Herbert Simon's Decision Making Theory*. Retrieved January 10, 2021, from Harappa Learning Prvate Limited: https://harappa.education/harappa-diaries/herbert-simons-decision-making-theory/

Herbert, S. A. (1947). *Administrative Behaviour: A Study of Decision Making Processes in Administrative Organizations*. Free Press.

Herbert, S. A. (1965). Administrative decision making. *Public Administration Review*.

Hern, A. (2022, May 12). *Planned EU rules to protect children online are attack on privacy, warn critics*. Retrieved June 1, 2022, from The Guardian: https://www.theguardian.com/society/2022/may/12/planned-eu-rules-to-protect-children-online-are-attack-on-privacy-warn-critics

Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS Explaind.* Routledge.

Hitt, M., & Collins, J. D. (2007). Hitt, M. A., & Collins, J. D. (2007). Business ethics, strategic decision making, and firm performance. *Business Horizons*.

Hoy, W. (2019). *Decision-Making Theory.* Retrieved Januar 7, 2022, from Wayne K. Hoy: https://www.waynekhoy.com/wp-content/uploads/2018/11/Theory-of-Decison-Making.pdf

Ipsos SoMe team. (2022, April 21). *Ipsos SoMe-tracker Q1'22.* Retrieved May 10, 2022, from Ipsos: https://www.ipsos.com/nb-no/ipsos-some-tracker-q122

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: antecedents and consequences. *Journal of Msrketing, 57*(3).

Jones, T. M. (1991, April 2). Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model. *Academy of Management Review*.

Jussupow, E., & Benbasat, I. (2020, June). Why are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion. *Information Systems*.

Kloppen, K., Haugland, S., Svedin, C. G., Mæhle, M., & Breivik, K. (2016). Prevalence of child sexual abuse in the Nordic countries: A literature review. *Journal of child sexual abuse*.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*.

Kripos. (2019). *Seksuell utnyttelse av barn og unge over internett.* Kripos.

Li, B. (2008). The Classical Model of Decision Making Has Been Accepted as not providing an Accurate Account of How People Typically Make Decisions. *International Journal of Business and management*(3).

LoBiondo-Wood, G., & Haber, J. (2014). *Nursing research: Methods and critical appraisal for evidence-based practice* (Vol. 8).

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm Appreciation: People Prefer Algorithmic To Human Judgment. *Organizational Behavior and Human Decision Processes*, pp. 90–103.

Madhavan, P., Wieggmann, D. A., & Lacson, F. C. (2006, February). Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids. *Human Factors The Journal of the Human Factors and, 48*(2), pp. 241-256.

Martin, E., & Moore, A. (2020, June). NEWS & VIEWS. *Tapping into artifcial intelligence: Advanced technology to prevent crime and support reentry.*

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing, 81*(1).

Medietilsynet. (2020). *Barn og medier 2020.* Medietilsynet.

Mildebrath, H. (2022, April). *Proporsal of regulation on a temporary dergation from certain provisions of the e-privacy directive of the purpose of combating child sexual abuse online/after 2020-3.* Retrieved from Legisative Train 04.2022: https://www.europarl.europa.eu/legislative-train/carriage/temporary-derogation-from-the-e-privacy-directive-for-ott-services/report?sid=

Muscanell, N. L., & Guadagno, R. E. (2021). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*.

Mæland, K. B. (2020, November 6). *Neste uke i Oslo tingrett blir overgrep mot barn og sosiale medier som Snapchat tema i flere straffesaker*. Retrieved May 10, 2022, from Nettavisen: https://www.nettavisen.no/nyheter/snapchat-og-overgrep-mot-barn-blir-tema-i-flere-rettssaker/s/12-95-3424042867

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3).

Newman, D. (2017, July 11). *Artificial Intelligence: To Be Feared or Embraced*. Retrieved April 17, 2022, from Forbes: https://www.forbes.com/sites/danielnewman/2017/07/11/artificial-intelligence-to-be-feared-or-embraced/?sh=1fdfa52f6f09

O'Brien, C. (2022, January 19). *How Do Social Media Algorithms Work?* Retrieved April 28, 2022, from Digital Marketing Institute: https://digitalmarketinginstitute.com/blog/how-do-social-media-algorithms-work

Onkal, D., Goodwin, P., & al., e. (2009, October). The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments. *Journal of Behavioral Decision Making, 22*(4), pp. 390-409.

Oracle. (2022, June 23). *What is Big Data?* Retrieved May 25, 2022, from OCI: https://www.oracle.com/big-data/what-is-big-data/

Oslo politidistrikt. (2022). *Anmeldt kriminalitet i Oslo politidistrik 2021.* Oslo politidistrikt.

Oxford Learner's Dictionaries. (n.d.). *Childe abuse*. Retrieved May 10, 2022, from Oxford Learner's Dictionaries: https://www.oxfordlearnersdictionaries.com/definition/english/child-abuse?q=+child+abuse+

Oxford Learner's Dictionaries. (n.d.). *Grooming*. Retrieved May 10, 2022, from Oxford Learner's Dictionaries: https://www.oxfordlearnersdictionaries.com/definition/english/grooming

Oxford Reference. (2020). *Artificial intelligence.* Retrieved May 20, 2022, from Oxford References: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960Oxford Ref

Plangger, K., & Montecchi, M. (2020). Thinking beyond privacy calculus: Investigating reactions to customer surveillance. *ournal of Interactive Marketing, 50*.

Prahl, A., & Swol, L. V. (2017, March). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Fprecasting*.

Quayle, E. (2020, December). Prevention, disruption and deterrence of online child sexual exploitation and abuse. *Era Forum* .

Robyn, D. M. (1979). The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist, 34*(7), pp. 571-582.

Russell, S., & Norvig, P. (2016). *Artificial intelligence: a modern approach.* (Vol. 3).

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences* 3(3), pp. 417-457.

Selart, M., & Johansen, S. (2011). Ethical decision making in organizations: the role of leadership stress. *Journal of business ethics*.

Sethi, D., Yon, Y., Parekh, N., Anderson, T., Huber, J., Rakoavc, I., & Meinck, F. (2018). *European status report on preventing child maltreatment.* World Health Organization.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019, July 13). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review, 4*(61), pp. 66-83.

Simonsen, J. (1994). *Administrative Behavior: How Organizations can be Understood in Terms of Decision Processes.*

Smith, A. (2020, April 8). *Using Artificial Intelligence and Algorithms ..* Retrieved May 25, 2022, from *Federal Trade Commission*: https://www.privacysecurityacademy.com/wp-content/uploads/2021/01/Using-Artificial-Intelligence-and-Algorithms-_-Federal-Trade-Commission.pdf

Srinivasan, R., & Sarial-Abi, G. (2021). When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors. *Journal of Marketing*.

Statista Research Department. (2022, May 5). *Average time spent on social media in Norway in 2019, by generation*. Retrieved May 29, 2022, from Statista: https://www.statista.com/statistics/1033973/time-spent-on-social-media-in-norway-by-generation/

Statistisk sentralbyrå. (2022, April 26). *Fakta om Internett og mobiltenefon*. Retrieved May 5, 2022, from Statistisk sentralbyrå Statistics Norway: https://www.ssb.no/teknologi-og-innovasjon/faktaside/internett-og-mobil

Sunde, N., & Sunde, I. (2021). Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse: Part I – The Theoretical and Technical Foundations for PrevBOT. *Universitetsforlaget*.

Sushama, C., Sunil Kumar, M., & Neelima, P. (2021). Privacy and security issues in the future: A social media. *Materials Today: Proceedings*.

Sylwander, K. R., Vervik, A.-K., & Greijer, S. (2021). *Online Child Sexual Exploitation and Abuse: a Review of Norwegian Case Law.* ECPAT Norway.

Whitley, B. (2002). *Principles of Research in Behavioral Science.* McGraw-Hill.

Zhang, C., & Sun, J. e. (2010, July). Privacy and Security for Online Social Networks: Challenges and Opportunities. *IEEE Network, 24*(4).

Terre des Hommes. (2022). *Update 2022: Terre des Hommes stops undercover and cyber activities*. Retrieved from Terre des Hommes : https://www.terredeshommes.nl/en/programs/sweetie

The Editors of Salem Press. (2016). *Decision Making & Crisis Managment.* Salem Press.

The Norwegian Ministry of Children and Families. (n.d.). *Protection of children against sexual exploitation and abuse.* The Norwegian Ministry of Children and Families.

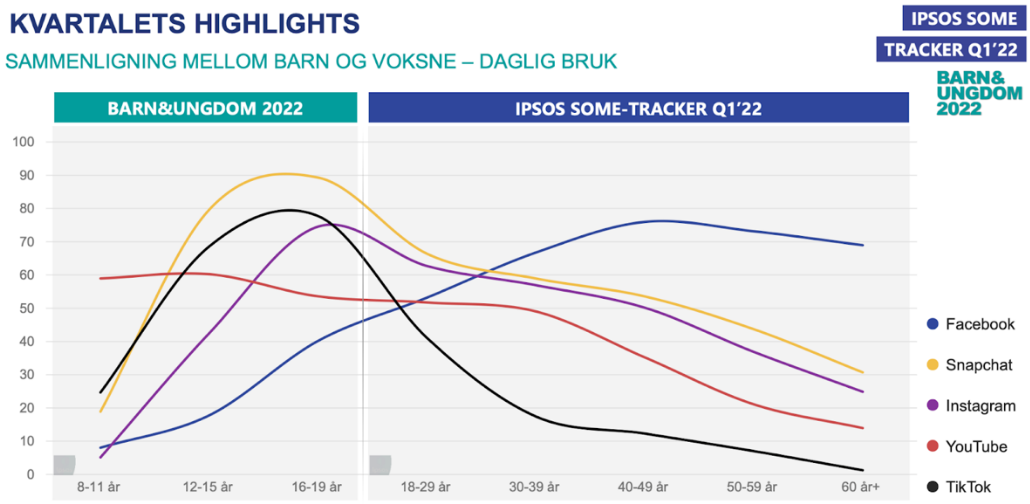
Tidey, A. (2022, May 11). *Here's how the EU's plan to tackle online child abuse could impact your privacy*. Retrieved May 13, 2022, from Euronews.: https://www.euronews.com/my-europe/2022/05/11/here-s-how-the-eu-s-plan-to-tackle-online-child-abuse-could-impact-your-privacy

UBS. (n.d.). *Herbert A. Simon*. Retrieved January 6, 2022, from UBS Nobel Perspetives: https://www.ubs.com/microsites/nobel-perspectives/en/laureates/herbert-simon.htm

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019, February 14). Making sense of recommendations. *Journal of behaviornal Decision Making*.

## 10.0 Appendix

**Appendix 1 -** Ipsos SoMe Team



Source: Ipsos SoMe team. (2022, April 21). Ipsos SoMe-tracker Q1'22. Retrieved May 10, 2022, from Ipsos: https://www.ipsos.com/nb-no/ipsos-some-tracker-q122

**Appendix 2** – The Survey

**Q1:** I confirm that I am over 18 years old, and accept that my response will be used for analysis in this master thesis.

- Yes
- No

**Q2:** Which of these social media platforms do you use regularly (more than 4 times a week)?

- Facebook
- Snapchat
- TikTok
- Instagram
- WhatsApp
- Messenger
- LinkedIn
- Others

**Q3:** In this question we refer to social media platforms as interactive technologies and digital channels that facilitate the creation and sharing of information, ideas, interests, and other forms of expression through virtual communities and networks.

How often do you use social media platforms to...

| | Never | Sometimes | About half the time | Most of the time | Always |
|---|---|---|---|---|---|
| send messages and/or pictures to family/friends | | | | | |
| follow influencers and celebrities | | | | | |
| follow friends and family | | | | | |
| be updated on trends and news | | | | | |
| meet/communicate with strangers | | | | | |
| work/school | | | | | |
| other reasons | | | | | |

**Q4:** How much time do you spend on social media platforms each day? (If you are unsure check your screen time on your phone)

- 1-2 hours
- 3-4 hours
- 5-6 hours
- More than 6 hours

**Q5:** In this question we refer to surveillance as the act of monitoring and logging your online data and traffic by a third party, such as the government, Internet service providers, Big Tech companies like social media companies, or criminals. How much do you agree with these assumptions:

|  | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| What I share with my friends/family/followers is private, and I would not like a third party to see it. |  |  |  |  |  |
| I would not like social media companies to surveillance my private chats. |  |  |  |  |  |
| I would feel that my privacy was invaded if a social media company was looking into my private chats. |  |  |  |  |  |
| Using social media platforms has made me feel anxious and afraid of being surveilled. |  |  |  |  |  |
| I always read the terms of privacy when joining a new social media platform. |  |  |  |  |  |

**Q6:**

The following questions are about the personal feeling of privacy:

|  | 1 (low degree) | 2 | 3 | 4 | 5 | 6 | 7 (high degree) |
|---|---|---|---|---|---|---|---|
| To which degree do you feel that social media companies protect your privacy? |  |  |  |  |  |  |  |
| To what degree have you experienced that a social media company has violated your privacy? (e.g. false profiles with your information, leak of private pictures, identity theft etc.) |  |  |  |  |  |  |  |
| To what degree would you allow social media companies to breach your privacy when they are suspicious of data being missused. |  |  |  |  |  |  |  |

**Q7:** I have felt judged by others based on something I have posted or shared on social media

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

*Artificial Intelligence (AI) is defined as the simulation of human intelligence processes by machines, especially computer systems.*

**Q8:** If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity like your messages and shared photos?

- AI robot
- Humans

**Q9:** The spread and sharing of illegal materials online is increasing. The European commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile?

- An AI robot
- A human

*Display question: If "An AI robot" is selected*

**Q10:** Since you chose AI robots instead of a human, please rate the following assumptions:

| | 1 (strongly disagree) | 2 | 3 | 4 | 5 | 6 | 7 (strongly agree) |
|---|---|---|---|---|---|---|---|
| I prefer an AI robot instead of a human because it feels like less of a threat to my privacy | | | | | | | |
| I prefer screening by an AI robot because it feels less personal since they don't have emotions and feelings | | | | | | | |
| I'm afraid that humans will judge me, and make decisions based on this judgment | | | | | | | |
| It feels less intimidating if an AI robot goes through my online activity than if a human does it | | | | | | | |
| I perceive AI robots to be nondiscriminatory (def. not making an unfair or prejudicial distinction between different categories of people or things) | | | | | | | |

*Display question: If "A human" is selected*

**Q11:** Since you chose a human instead of AI robots, please rate the following assumptions:

| | 1 (strongly disagree) | 2 | 3 | 4 | 5 | 6 | 7 (strongly agree) |
|---|---|---|---|---|---|---|---|
| I prefer a human instead of an AI robot because it feels like less of a threat to my privacy | | | | | | | |
| I prefer screening by a human because they have the best ability to make rational decisions | | | | | | | |
| I am afraid that AI is discriminatory, and that it will misjudge me | | | | | | | |
| It feels less intimidating if a human goes through my online activity than if an AI robot does it | | | | | | | |
| I perceive humans to be nondiscriminatory (def. not making an unfair or prejudicial distinction between different categories of people or things) | | | | | | | |

**Q12:** Do you think AI robots or humans are the least discriminatory?

- AI robots
- Humans

*Display question: If "AI robot" is selected*

**Q13:** Why do you think AI robots are the least discriminatory?

- Because AI robots don't have feelings and emotions
- Because AI robots don't have "human judgement"
- Because humans have a gut feeling
- Because AI robots are made for solving concrete tasks
- Because AI robots are more suitable for making rational decisions

*Display question: If "Humans" is selected*

**Q14:** Why do you think humans are the least discriminatory?

- Because humans have feelings and emotions
- Because humans have "human judgement"
- Because AI robots don't have a gut feeling
- Because humans can see thing from different points of view
- Because humans are more suitable for making rational decisions

**Q15:** Do you trust AI?

- Definitely not
- Probably not
- Might or might not
- Probably yes
- Definitely yes

**Q16:** How much do you agree with these assumptions:

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| It makes me anxious to think about another human reading my personal messages. | | | | | |
| It makes me anxious to think about an AI robot reading my personal messages. | | | | | |

**Q17:** I would allow social media platforms to look through my online activity if it is for the betterment of society (e.g. to avoid crime, child abuse, hate or discrimination).

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

**Q18:** In which of the following examples would you be willing to accept social media platforms to surveil your online activity?

- For commercial purposes
- For crime prevention causes
- To prevent child abuse
- To provide personalized offers and ads
- To provide followers suggestions
- To prevent violence, hate and discrimination
- To prevent identity-theft

**Q19:** Imagine a scenario where an AI robot is screening the messages between you and a friend, for an investigation of a case regarding child abuse. How anxious would you be about being misinterpreted by the AI robot? (knowing that you are innocent)

- Extremely anxious
- Somewhat axious
- Neither anxious nor unanxious
- Somewhat unaxious
- Extremely unaxious

**Q20:** How old are you?

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65-74
- 75-84
- 85 or older

**Q21:** How do you describe yourself?

- Male
- Female
- Non-binary / thrid gender
- Prefer not to say

## **Appendix 3 –** Reliability Statistics

**Reliability Statistics**

| Cronbach´s Alpha | Cronbach´s Alpha Based on Standardized Items | N of Items |
|---|---|---|
| .858 | .896 | 17 |

## **Appendix 4 –** Classification table Scenario 1 without moderators

Classification Table[a]

| | | *Predicted* | | |
|---|---|---|---|---|
| | | If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity like your messages and shared photos? | | |
| *Observed* | | AI robot | Human | Percentage Correct |
| If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity like your messages and shared photos? | AI robot | 79 | 1 | 98.8 |
| | Human | 14 | 9 | 39.1 |
| Overall Percentage | | | | 85.4 |

a. The cut value is .500

## **Appendix 5 –** Linear regression showing IV 2

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 2 | Fear of being misunderstood by AI | -.196 | .204 | .922 | 1 | .337 | .822 |
| | *Constant* | -.592 | .649 | .833 | 1 | .361 | .553 |

## **Appendix 6 –** Linear regression showing IV 4

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 4 | Fear of breach of privacy | .123 | .188 | .429 | 1 | .512 | 1.131 |
| | *Constant* | -1.646 | .705 | 5.456 | 1 | .019 | .193 |

## **Appendix 7 –** Linear regression showing IV 5

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 5 | Anxiety of being surveilled | .265 | .213 | 1.542 | 1 | .214 | 1.978 |
| | *Constant* | -1.892 | .656 | 8.324 | 1 | .004 | .151 |

## **Appendix 8 –** Classification table Scenario 1 with moderators

| Classification Table[a] | | | | |
|---|---|---|---|---|
| | | *Predicted* | | |
| | | If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity like your messages and shared photos? | | |
| *Observed* | | AI robot | Human | Percentage Correct |
| If a social media company should surveil your online activity, would you prefer an AI robot or a human to go through your online activity like your messages and shared photos? | AI robot | 79 | 1 | 98.8 |
| | Human | 13 | 10 | 43.5 |
| Overall Percentage | | | | 86.4 |

a. The cut value is .500

## **Appendix 9 –** Multiple regression on IV 5 with the moderator "The purpose"

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| IV 5 | Anxiety of being surveilled | .273 | .239 | 1.308 | 1 | .253 | 1.314 |
| The moderator "The purpose of surveillance" | | | | | | | |
| 1 | For commercial purposes | 1.465 | 1.080 | 1.839 | 1 | .175 | 4.328 |
| 2 | For crime prevention causes | -.272 | .679 | .160 | 1 | .689 | .762 |
| 3 | To prevent child abuse | -1.102 | .880 | 1.568 | 1 | .210 | .332 |
| 4 | To provide personalized offers and ads | 1.040 | .744 | 1.953 | 1 | .162 | 2.830 |
| 5 | To provide followers suggestions | .814 | .957 | .725 | 1 | .395 | 2.258 |
| 6 | To prevent violence, hate and discrimination | .568 | .742 | .587 | 1 | .443 | 1.765 |
| 7 | To prevent identity-theft | -.359 | .697 | .265 | 1 | .607 | .699 |
| | *Constant* | -1.892 | .656 | 8.324 | 1 | .004 | .151 |

## **Appendix 10 –** Classification table Scenario 2 without moderators

| Classification Table[a] | | | | |
|---|---|---|---|---|
| | | *Predicted* | | |
| | | The spread and sharing of illegal materials online is increasing. The European commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile? | | |
| *Observed* | | AI robot | Human | Percentage Correct |
| The spread and sharing of illegal materials online is increasing. The European commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile? | AI robot | 67 | 2 | 97.1 |
| | Human | 23 | 12 | 34.3 |
| Overall Percentage | | | | 76.0 |

a. The cut value is .500

## **Appendix 11 -** Multiple regression with moderators

| IV | | B | S.E | Wald | df | Sig./p | Exp(B) |
|---|---|---|---|---|---|---|---|
| 1 | Trust in AI | -.458 | .331 | 1.915 | 1 | .166 | .632 |
| 2 | Fear of being misunderstood by AI | .059 | .206 | .085 | 1 | .774 | 1.061 |
| 3 | Fear of discrimination by AI and humans | 3.435 | 1.127 | 9.290 | 1 | .002 | 31.045 |
| 4 | Fear of breach of privacy | .215 | .200 | 1.145 | 1 | .285 | 1.239 |
| 5 | Anxiety of being surveilled | -.067 | .219 | .093 | 1 | .761 | .936 |
| Moderators | | | | | | | |
| 1 | Age | .035 | .208 | .029 | 1 | .865 | .036 |
| 2 | Gender | .0538 | .361 | 2.228 | 1 | .136 | 1.713 |
| | *Constant* | -4.751 | 2.035 | 5.452 | 1 | .020 | .009 |

## **Appendix 12 –** Classification table Scenario 2 with moderators

Classification Table[a]

| | | *Predicted* | | |
|---|---|---|---|---|
| | | The spread and sharing of illegal materials online is increasing. The European commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile? | | |
| *Observed* | | AI robot | Human | Percentage Correct |
| The spread and sharing of illegal materials online is increasing. The European commission and the Norwegian government have proposed that telecom providers (like Telenor) should screen private messages, in an attempt to prevent online abuse of children. If this policy is implemented, who would you be most comfortable with screening your profile? | AI robot | 67 | 2 | 97.1 |
| | Human | 22 | 10 | 28.6 |
| Overall Percentage | | | | 74.0 |

a. The cut value is .500