FEAR NOT, HERE IS THE BOT

AN EXPERIMENTAL STUDY OF HOW CONSUMERS RATE SATISFACTION AND PERFORMANCE IN SERVICE ENCOUNTERS WITH CHATBOTS AND EMPLOYEES

ARVID BERGSTRAND

YOUCEF DJEHICHE

Bachelor Thesis
Stockholm School of Economics
2022

Fear not, here is the bot – An experimental study of how consumers rate satisfaction and performance in service encounters with chatbots and employees

Abstract:

Artificial intelligence is entering e-commerce customer service in force. Chatbots, conversational agents employed to address consumers' needs and questions, are steadily replacing employees, a development claimed to increase the quality of service. This quantitative study examines the levels of customer satisfaction, and perceived employee performance, consumers experience after having interacted with either a chatbot, or a human service employee. The purpose is to provide some further insight into when, and for what demographic groups of customers, it is appropriate to employ chatbots rather than humans in an e-commerce chat service setting. The results suggest that in this specific case, there are no differences in customer satisfaction nor perceived employee performance, and thus no human bias for or against chatbots. Based on these results a suggestion directed at e-commerce actors is to continue investing in and employing chatbots, to focus on outcome rather than how that outcome is delivered. Replacing employees can provide substantial cost saving opportunities, and perhaps increase service quality.

Keywords:

Algorithm aversion, Algorithm appreciation, Customer service encounter, Customer service, AI, Algorithm, Chatbot, E-commerce

Authors:

Arvid Bergstrand (24806)
Youcef Djehiche (24876)

Tutors:

Hanna Berg, Research Fellow, Department of Marketing and Strategy

Examiner:

Patric Andersson, Associate Professor, Department of Marketing and Strategy

Bachelor Thesis

Bachelor Program in Business & Economics
Stockholm School of Economics

Acknowledgements

We would like to thank Magnus Söderlund, who has given us input and insights along this journey, Patric Andersson, who has held the entire course together and made sure we have got all the knowledge needed, and all of those who responded to our survey. Beyond those we want to send a special thanks to Hanna Berg. Not only has she guided us in our research, this thesis, and provided us with invaluable knowledge, but she has also helped us by encouraging and keeping us motivated throughout the writing of this thesis. Thank you!

Definitions

AI: Artificial intelligence, i.e., non-biological intelligence (Ostrom et al., 2018).

Algorithm appreciation: “Positive behavior and attitudes towards the algorithm compared to a human agent” (Jussupow et al., 2020, p. 4).

Algorithm aversion: “A biased assessment of an algorithm which manifests in negative behaviors and attitudes towards the algorithm compared to a human agent” (Jussupow et al., 2020, p. 4).

Attribution bias: The tendency to quickly form judgments drawn from personal beliefs rather than the actual situation at hand.

Chatbots: Conversational tools employed to converse, address, and handle a variety of customer needs and requests, most commonly through text (Crolic et al., 2022).

Customer satisfaction: Will in this study refer to how happy a customer is with a service interaction.

Customer service agent: Will in this study refer to a human who works in customer service.

E-commerce: Electronic commerce, i.e. electronically conducted commercial transactions.

Perceived employee performance: Will in this study refer to the customer's evaluation of employee service behaviors.

Service AI: “The configuration of technology to provide value in the internal and external service environments through flexible adaptation enabled by sensing, learning, decision-making and actions” (Bock et al., 2020, p. 1).

Technophobia: The fear of or aversion to advanced technology and complicated products, specifically computers.

The fourth industrial revolution: An intermix of different technologies combined with the merging of digital, physical and biological spheres (Schwab, 2016).

The service encounter: The interaction and direct contact between a consumer and the service provider.

Contents

1. Introduction

As humans, we are from birth dependent on, and throughout all aspects of life interact with other human beings (Epley, 2018). A new area of research has emerged, as many of the interactions previously between humans, such as those within e-commerce customer service have been automated (Crolic et al., 2022). Technological advancements are constantly shifting and disrupting industries. The digital era, which is commonly known as the third industrial revolution, is arguably slowly coming towards its end. The next big industrial revolution is at its dawn: the era of automation (Scwab, 2016).

One interesting aspect of this evolution is the human interaction with non-human intelligence. As human service employees are increasingly being replaced by bots and algorithms, the industries and businesses that utilize these technologies are faced by new challenges, and questions (Crolic et al., 2022). Is there a tendency to rate interactions with humans as more positive than chatbots in digital settings? Among what demographic groups do such tendencies exist? This thesis aims to look at a narrow aspect of these questions, within an e-commerce service setting, by comparing customer satisfaction and perceived employee performance after having read the same conversation with either a human service employee or a chatbot. Online customer service is one of the areas that were quickest to utilize this new technology and where the development has gotten furthest (Belanche et al., 2020; Lu et al., 2020), which explains why the above posed questions play an ever more important role.

1.1 Background

1.1.1 Background on e-commerce

Following the internet's breakthrough during the 1970's and 1980's the first website was published in 1991 (Nix, 2018). Shortly thereafter, in 1993, the first browser to access the internet was introduced. Today, e-commerce has dramatically changed the setting of business-to-consumer sales and transactions. Commerce has largely moved from the traditional brick-and-mortar stores to electronic marketplaces and internet-based supply chains (Zwass, 2019).

The e-commerce market has grown to a significant size, and during the pandemic of 2020 and 2021 it skyrocketed. Changes in demand and consumer behavior were reflected in consumers turning to their devices for a broader range of purchases than before (Phaneuf, 2022). Worldwide retail e-commerce sales soared from 1.336 billion U.S dollars in 2014 to 4.938 billion U.S dollars in 2021 (Chevalier, 2022), almost a fourfold increase. Although the levels are expected to stabilize during 2022, sales will likely exceed 5 billion U.S dollars (Cramer-Flood, 2022), and some expect it to continue growing by around 50% come 2025, breaching 7.4 billion U.S dollars (Chevalier, 2022). In Sweden, e-commerce sales reached a peak of around 330 billion SEK in 2019 but decreased to around 215 billion SEK during 2020 because of the pandemic. (Salesforce, 2021)

1.1.2. Background on AI and chatbots

Artificial intelligence, AI, is on the rise in all aspects of society, not the least within marketing and e-commerce, where companies are implementing AI-driven tools to improve customer experiences (Crolic et al., 2022). Conversational agents are employed to address and handle a variety of customer needs and requests, most commonly through text. These tools are called chatbots.

AI is defined by Syam & Sharma (2018, p. 136) as “the ability of machines to mimic intelligent human behavior”, specifically, “cognitive functions that we associate with the human mind, including problem-solving and learning”, or by Ostrom et al. (2018, p. 80) as “non-biological intelligence”. One should, however, note that AI has the capacity to far exceed human capabilities, why these definitions may be limiting (Bock et al., 2020). Rather, Bock et al. (2020, p. 1) claim that service AI should be defined as “the configuration of technology to provide value in the internal and external service environments through flexible adaptation enabled by sensing, learning, decision-making and actions”.

Receiving service from a bot rather than a human service agent has become increasingly frequent (Belanche et al., 2020; Lu et al., 2020), and several observers are confident that the marketplace and workforce for services will be altered (Broadbent, 2017; Murphy et al., 2019), improving quality while reducing costs (De, 2018). A trend that is expected to continuously accelerate going forward (Mende et al., 2019; Söderlund, 2021). AI is claimed to be able to outperform humans, especially in simple and repetitive tasks (Huang & Rust, 2018; Xu et al., 2020). It is thus hypothesized that consumers’ steadily increasing demand for convenience and speed has led to increased acceptance of these types of self-service technologies (Collier & Kimes, 2013; Grewal et al., 2017).

As will be returned to further, there is a rift in the research on humans and algorithms. The major consensus, however, is in the area known as algorithm aversion, which claims that users prefer humans over algorithms even in situations where the algorithm has been proven to be superior (Jussupow et al., 2020). Contradictorily, industries are steadily increasing the investments in, and use of, bots and algorithms, not the least within customer service (Belanche et al., 2020; Lu et al., 2020). The area of algorithms is moving very quickly (Mende et al., 2019; Söderlund, 2021), and this may be a sign that practice is ahead of theory.

The consequences of implementing chatbot and AI technologies are prevalent. As AI and chatbot technology is not at the level where it can excel at tasks other than general and repetitive assignments, the tool often struggles at more complex tasks. This has made brought consequences for customers and business who find the tool as less useful than a human employee, thus creating a negative narrative regarding chatbots. There is also a general negative connotation around chatbots, and specifically surrounding the terminology “bot”. The term “bot” is generally attributed to programs used to cause some sort of havoc in digital systems, a tool often used by hackers. Chatbots on the other hand have not historically been used for illegitimate purposes, as the technology is mostly used as a conversation tool. (Oracle, 2022) Further, as the technology is dependent on how experienced and trained the tool is, the level that different chatbots operate in varies. Thus, the opinion surrounding the technology, from business and consumer sides, varies greatly depending on how good the tool they are exposed to is. (Deloitte, 2018)

1.1.3. The future of AI and chatbots

Many claim that we are stepping into a new era of technological advancements. The fourth industrial revolution (Bock et al., 2020), characterized by an intermix of different technologies combined with the merging of digital, physical and biological spheres (Scwab, 2016). Others predict that the future will be entirely led by automation and AI, and some predict that by 2025 95% of customer interactions, including live telephone calls and chat conversations, will be handled by AI (Marshall, 2017). Furthermore, some believe that already at that point, consumers will be unable to ‘spot the bot’, the conversational agents will be so human-like that we will not be able to tell the difference between them, and a person (Marshall, 2017).

Although these predictions could be claimed to be outdated since these technologies are developing very quickly, more recent figures seem to confirm earlier predictions. By 2024 consumer retail spending via chatbots are anticipated to reach 142 billion U.S dollars, up from only 2.8 billion U.S dollars in 2019, a fifty-fold increase in five years

(Insider, 2022). Since these chatbots can simulate natural language processing and actively collect data on it, each interaction improves their abilities (Insider, 2022).

With this development there is concern that the algorithms will exert too much power over our lives, as discussed by Sumpter (2018), who claim that algorithms and AI are more than they seem to be. As their abilities to solve problems improve, so will they also be able to gather more data on us, and better predict our needs and requests. A development which may be good, considering that the level of service will be improved, but which also may be troubling as this knowledge may be used to influence us, as discussed by Sumpter (2018).

1.2 Problem formulation

The use of AI and chatbots in customer service is evidently widespread and will likely increase at an even higher rate in the future. The technology is steadily developing and becoming more and more human-like (Marshall, 2017). What is less clear, however, is how consumers perceive the interaction with conversational tools, in comparison with human customer service agents. With what levels of customer satisfaction and perceived employee performance do the different approaches to the service encounter respectively leave customers? Previous conclusions on the subject differ widely and often contradict each other directly. Apart from algorithm aversion, another strain of theory is claiming the very opposite, that human sentiments towards chatbots are more positive than towards human service employees (Jussupow et al., 2020).

There are advantages with the use of AI, such as quicker response times and higher availability (Huang & Rust, 2018; Xu et al., 2020). Our objective, however, is to look at a scenario where all these external circumstances are stripped away, and determine what differences there are after having conversed with either a chatbot or a human service employee. Is algorithm aversion or algorithm appreciation present among consumers in an e-commerce service encounter?

1.3 Research purpose and research question

The purpose of the thesis is to examine how consumers perceive service encounters from chatbots and employees. In particular, we examine whether consumers will rate customer satisfaction and perceived employee performance differently when having been in contact with a chatbot and a human. We further strive to conclude what potential differences, if any, there are in the responses of different demographic groups, divided on age, education, and gender.

In short, we strive to conclude whether there are human biases among consumers towards algorithms and AI conversational agents, and whether these impact perceived employee performance and customer satisfaction after a conversation. Based on this purpose the research question we aim to answer is therefore the following:

To what extent do consumers rate human-to-human-interaction in e-commerce customer encounters as more positively than interacting with a chatbot? What factors explain this potential difference? To what extent could demographic variables moderate this difference?

1.4 Delimitations

As this is a bachelor thesis, there are formal requirements, as well as resource and time restraints, that we have taken into consideration during the creation of it. The consequences of which are the following delimitations.

A delimitation is that we were not able to set up an actual conversation between each participant of the study and a chatbot or human agent, likely reflected in a lower degree of realisticity. Due to limited resources, we supplied the participants with a pre-written scenario, a conversation where they had no possibility to control either side of the conversation, and likewise were not able to consider the speed of replies and other factors. This choice was partly deliberate, to reduce the number of differences in the cases. Nonetheless, this does limit the level of applicability and generalizability in the conclusions.

Another limitation is that our study is a snapshot in time. With greater resources and time, it would have been interesting to study whether individual attitudes change over time as the technology evolves and becomes more common.

1.5 Expected contributions

Our expected contribution is to provide clarity on human biases in this specific situation and implications on how best to design and adapt customer service in e-commerce. Our results, and the implications and generalizability of them, will be very limited because of the specificness of the case. However, despite the possible limitedness in our implications, they will contribute to the ever-growing accumulated knowledge of in which situations humans are biased for or against algorithms, and AI. As the scope of service AI continues to grow, there is significant value in knowing in which situation, and in interaction with what demographic groups it is, or is not, beneficial to use chatbots. By doing so we also expect to be able to determine whether businesses who are investing heavily in chatbots are moving in the right direction.

2. Literature review and theoretical framework

As a basis for the literature search and the theoretical foundation of this thesis we have turned to Magnus Söderlund’s work, mainly on the service encounter and customer satisfaction, a field in which he has published several studies. Beyond Söderlund, we have used databases and search tools such as Scopus and Google Scholar to search for other sources of information, previous studies, papers, and research.

2.1 Previous work

While there is a lot of theory, and plenty of previous studies which we have been influenced by and used, there are, to our knowledge, no other studies which have tested this specific case. While the study by Li et al. (2020) had some similarities, their sample of respondents were entirely Chinese, a group of people who are significantly more accustomed to automation, the effects of which can be compared to age and gender differences which we have discussed above. Furthermore, their study did not include any comparisons between humans and algorithms, and they did neither focus entirely on chatbots.

As has been mentioned, we have based a lot of our work on Söderlund’s research. A direct example is the usage of perceived employee performance in this study, which is a measurement that originates in Söderlund’s (2018) studies. Beyond that, we have also used Fornell’s (1992) questions and measurement scales for customer satisfaction. The usage of Fornell’s (1992) questions is a reliable choice, as the questions for measuring customer satisfaction are widely used and generally accepted.

As utilized by Söderlund (2021), a between-subjects experiment, simulating two different versions of an interaction or experience in which an independent factor which the authors wishes to test the effects of are manipulated are distributed to two groups of test subjects or respondents, or similar quantitative methods are common within this research area.

While systematic variability, conscious variability aimed at different needs and customers such as offering solutions that are generally appreciated by a demographic group, can result in positive effects for the company, unsystematic variability, unconscious variability caused by for example different quality among service employees resulting in some customers receiving worse service, should generally be avoided as it results in lower satisfaction among customers.

2.2 The service encounter

There is an abundance of research on the service encounter, which has shown the importance of a consumer interacting with other humans. The employee has a great impact on customer satisfaction. From the customer's point of view, the employee is the service, and therefore also the company (Bitner et al., 1990). Hence, the employee in a human-to-human service encounter has a direct effect on brand image, satisfaction with- and evaluations of the company (Bitner et al., 1990; Rafaeli & Pratt, 1993; Söderlund, 2012). This is reflected in the evaluation of the service and the entire company being heavily dependent on the service provider (Söderlund, 2021). An explanation of this reaction can be found within the affect infusion model (Forgas, 1995). The model claims that emotions awoken by a certain object, the employee in this case, can impact the evaluation of another, the company in this case.

Customers react differently to different service environments, and encounters, depending on their experienced level of control (Söderlund, 2012). An encounter which offers a low degree of experienced control results in negative feelings among customers, associated with the wish of leaving the interaction altogether. On the opposite, an encounter in which the customer experiences a high level of control is reflected in positive emotions and wanting to stay in the interaction (Foxall & Greenley, 1999). In line with the affect infusion theory, the positive emotions created in the interaction are further transferred to a positive attitude towards the company represented (Ward & Barnes, 2001). Contrarily, Ridgeway (1987) showed that customers are willing to give up control in the interaction if they believe it will result in the task being solved in a better way.

Human-to-human service encounters will by default vary in one way or another, known as service variability or service heterogeneity (Hoffman & Bateson, 1997). Variation generally comes in two forms, systematic and unsystematic. While systematic variability can be conscious and result in positive effects for the company, unsystematic variability in the service encounter should generally be avoided as it results in lower satisfaction among customers (Söderlund, 2012).

2.3 Customer satisfaction

Customer satisfaction is an evaluative judgment made by the customer after purchasing or consuming an offer, or being in contact with a company (Kotler, 1994). The satisfaction of a service encounter can be explained by a function of the match between the customer's expectations of the service encounter prior to the interaction and the perceived performance of the company, service employee, or chatbot, after the interaction (Kotler, 1994).

Söderlund (2018) further found that customer satisfaction is positively correlated with perceived employee performance, the customer's experience of the service quality of an interaction which often has a mediating role to customer satisfaction. Moreover, customer satisfaction is positively correlated to customer loyalty (Jones, 1996). To increase loyalty among customers one must thus increase customer satisfaction. Loyalty, in turn, is positively correlated to repurchase intentions, why customer satisfaction is very relevant for all types of customer interactions and service encounters.

2.4 Algorithm aversion

Opposing conclusions have been made, but the consensus is that the choice between a human and an algorithm is not based on rational, objective criterias. Rather, customers prefer humans even in situations where the algorithm has been proven to be superior (Jussupow et al., 2020). Within the service encounter, this should be reflected in more positive evaluations of human service employees than chatbots.

Algorithm aversion is defined by Jussupow et al. (2020, p.4) as a "biased assessment of an algorithm which manifests in negative behaviors and attitudes towards the algorithm compared to a human agent". However, Jussupow et al. (2020) also concluded that there is no clear conceptualization of algorithm aversion and that it thus is difficult to extract precise theory regarding whether users develop aversion towards algorithms.

Algorithm aversion consists of a biased assessment of the algorithm that is not implemented on human agents; reactions thus differ between the two (Jussupow et al., 2020), which should be reflected in different levels of customer satisfaction and perceived employee performance in the service encounter. Possible explanations to the differing reactions can be lower levels of trust in the algorithm (Madhavan & Wiegmann, 2007; Önkal et al., 2009), experienced appropriateness of the agent's or algorithm's decisions (Bigman & Gray, 2018; Palmeira & Spassova, 2015), or lower experienced authenticity of the algorithm's actions (Jago, 2019).

Social distance affects how others are evaluated and how we process information in interactions (Trope & Liberman, 2010). This could be an explanation as to why users would be averse towards algorithms, which are more socially distant than another human. Another explanation is provided by Luo et al. (2019) who concluded that, in the case of active salesbots, a situation similar to the service encounter, customers perceive bots as less knowledgeable, and less empathetic. Because of this, customers purchased less. Luo et al. (2019 p. 1) explained this as a "subjective human perception against machines" which can be equated to algorithm aversion.

Most of the research on algorithm aversion has not focused specifically on the service encounter, and no research has tested, or compared, either customer satisfaction or perceived employee performance. However, we believe that the results are applicable in the setting of this study. Chatbots are algorithms, and as these effects have been found in other situations, so should they be present in the service encounter.

2.5 Person positivity bias

An aspect that may play a role in the negative attitudes towards algorithms within the service encounter is perceived humanness. Humans often have a positive, rather than negative, attitude towards other humans (Sears, 1983). Person positivity bias is estimated to be a result of the inherently social aspects of being human and the promises of social connection, belongingness and intimacy that follows with other humans (Söderlund, 2016). Further, perceived similarity has been shown to have a positive impact on evaluations (Cialdini, 2007). Likewise, perceived humanness has a positive impact on trust (Castelo et al., 2019; Hadi, 2019), which in turn has a positive impact on evaluations within the service encounter, and customer satisfaction (Anderson & Narus, 1990).

It is important to mention that humans tend to ascribe human attributes to nonhuman things, so called anthropomorphization (Söderlund, 2021). This means that given the setting in which our study takes place, it is not necessary that the human agent will be experienced to have greater humanness than the chatbot. Regardless, person positivity bias is reflected in higher experienced humanness, which correlates with higher customer satisfaction and may thus be an explaining factor behind algorithm aversion within the service encounter.

2.6 Attribution theory

Mozafari et al. (2021) connected algorithm aversion to attribution bias, the tendency to quickly form judgments drawn from personal beliefs rather than the situation at hand. In a service encounter, consumers would depend on their negative perceptions towards chatbots. Li et al. (2020) came to similar conclusions and stated that customer's negative emotions towards chatbots are determined by subjective perceptions rather than objective, rational, facts.

Attribution theory, the causal explanations to questions such as "Why did this happen?", originally stems from social psychology works by Heider (1958). The explanations are based on beliefs, motivations, available information, the evaluation of the situation and the conditions around it. As claimed by Belanche et al. (2020), customers expect the outcome of an interaction with a bot in the service encounter to not yet be predictable or stable, in contrast to that of an interaction with a service employee, who they expect to be recruited through quality procedures and properly trained. Apart from what has previously been discussed, this provides a thorough explanation to algorithm aversion. However, one should keep in mind that the technical advancements on AI development are moving quickly, and so is the general perception of AI (Collier & Kimes, 2013; Grewal et al., 2017).

The studies discussed have not specifically tested either the effects on customer satisfaction, or perceived employee performance. However, as we have mentioned and discussed, there is widespread theory on the human bias for humans in contrast to algorithms in the service encounter. This should be reflected in customer satisfaction and perceived employee performance. These effects on the evaluation of the service encounter can be explained by algorithm aversion, person positivity bias and the attribution theory (Cialdini, 2007; Jussupow et al., 2020; Mozafari et al., 2021). Based on this we believe that we will have similar results and theorize the following;

H1a: Consumers faced with human service employee will be more satisfied than consumers faced with chatbots.

H1b: Consumers faced with human service employee will perceive employee performance to be greater than consumers faced with chatbots

2.7 Algorithm appreciation

Contrasting conclusions on the relation between humans and algorithms have, however, been made. Jussupow et al. (2020, p. 4) defined algorithm appreciation as a "positive behaviour and attitudes towards the algorithm", which stands in direct contrast to algorithm aversion.

Algorithm appreciation has been proven to exist, at least under certain criteria, by Logg et al. (2019) who showed that people prefer algorithms to other humans. More specifically, humans are prone to take the advice of algorithms above those of other humans and put greater trust in the judgment of algorithms on a wide range of different scenarios. Although the advice was identical, greater trust was put into that of the algorithm, even above the judgment of the participants themselves.

Similarly, in the context of service encounters, Tran et al. (2021) found that sentiments towards chatbots were less negative than those towards human agents, and that the sentiment towards human agents become more negative once a retailer implements a chatbot. Although the results differed somewhat depending on the retail sector, telco as compared to online fashion in this case, they found that consumers feel more positive towards chatbots than the human agents. Though Tran et al. (2021) did not use the term algorithm appreciation, the results substantiate the conclusions made by Logg et al. (2019) and show that such results exist in a chat-service encounter.

2.8 High and low task complexity

Other studies on algorithm aversion towards chatbots in the service encounter have found that sentiments may change depending on the degree of complexity in the task at hand. Xu et al. (2020) showed that AI is perceived to have greater problem-solving capacities and that customers show greater intent of usage of it when it comes to low complexity tasks. On the contrary, humans are perceived to have greater problem-solving abilities, and customers show greater intent of usage of human agents when faced with a high complexity task. Likewise, Mozafari et al. (2021) concluded that consumers generally are skeptical towards chatbots, but that they do trust them for services and tasks with low criticality. Confronted with high criticality needs, however, consumers do not trust the chatbot.

2.9 Attitudes towards AI based on age

On a broader scope than specifically AI and the service encounter, there is a lot of evidence that older people are more prone to avoid technology, commonly known as the digital divide between generations (Neves et al., 2018; Berkowsky et al., 2015). Technology avoidance and technophobia, generally defined as the fear of, and aversion towards, advanced technology, is prevalent in older populations. Among other things, this is characterized by lower levels of adoption of digital communication technologies, such as chatbots (Barbosa Neves et al., 2018; Berkowsky et al., 2015).

Literature generally depicts older age groups as technology resistant and non-users (Neves et al., 2018; Vines et al., 2015). A possible reason for this divide between generations is that older users were introduced to new technology at a, in relative terms, more developed age. This makes them less comfortable using them than younger age groups, such as the generation Z, so called digital natives who have grown up with technology (Cameron et al., 2001; Jin-Jong, 2015; Nimrod, 2018).

Christy et al. (2019) concluded, more narrowly within the service encounter, that older individuals express more avoidance toward automated communication, as compared to younger age groups, in e-health contexts. Furthermore, Li et al. (2020) showed that consumers' acceptance of receiving customer service via AI tools decreases with age. They claim that young consumers who are more used to, and more often exposed to, modern technologies show higher degrees of acceptance towards adopting, and using, new service methods. Older consumers on the other hand, are less accepting of trying, and using, new methods such as AI for their service enquiries.

These studies highlighting older generations' aversion towards complex technology in general, and advanced communication tools specifically, have not tested customer satisfaction or perceived employee performance. Neither have they solely looked at the service encounter. However, we believe that previous results should be reflected in these factors and further theorize;

H2a: Age moderates customer satisfaction. In comparison with younger age groups, older participants who have interacted with a human service employee will to a greater degree report higher customer satisfaction in contrast to those having interacted with a chatbot.

H2b: Age moderates perceived employee performance. In comparison with younger age groups, older participants that have interacted with a human service employee will to a greater degree report higher perceive employee performance in contrast to those having interacted with a chatbot.

2.10 Attitudes towards AI based on gender

Much like the generational divide, there is a lot of evidence and discussion on and around there being a “gender digital divide”. The term has been connected to differences, such as socioeconomic dissimilarities and inequalities, between those who have possibilities and abilities to utilize digital resources, and those who do not (Sheikh & Abbas, 2015). Women, at least historically, being an example of the latter.

Early studies, from the 80’s and 90’s, concluded that differences in technophobia between males and females do exist, with women being more technophobic. More recently and in line with earlier studies, Anthony et al. (2000) concluded that, depending on the specific technology’s diffusion, women do exhibit higher degrees of technophobia.

Other studies have come to similar conclusions in other aspects than specifically technophobia and the aversion towards advanced technology. Women show greater degrees of computer anxiety (Gilbert et al., 2003), and are less in favor of automatic cars (Hudson et al., 2019), for example. The structuration theory established that social structures influence behavior and thought paths around things such as technology, and AI (Anthony Giddens, 2013; Giddens, 1984). An example of which is chatbots within the service encounter.

While neither of these studies have tested customer satisfaction nor perceived employee performance between genders, and neither have focused on chatbots, there is a theoretical foundation which should be applicable in the service encounter. Based on this theoretical foundation, as well as the more historical results in nascent fields and studies, we theorize;

H3a: Gender moderates customer satisfaction. In comparison with male respondents, female participants who have interacted with a human service employee will to a greater degree report higher customer satisfaction in contrast to those having interacted with a chatbot.

H3b: Gender moderates perceived employee performance. In comparison with male respondents, female participants who have interacted with a human service employee will to a greater degree report higher perceived employee performance in contrast to those having interacted with a chatbot.

2.11 Attitudes towards AI based on education

In comparison to age, the literature on educational level and technophobia is far more limited. However, in plenty of studies on the generational divide, there have been additional findings on how educational levels impact the degree of technophobia. They have shown that there is a negative correlation between higher education and higher degrees of technophobia (Marescotti et al., 2021; Nimrod, 2018). The more educated consumers are, the less likely they are to be technophobic. Those with higher educational levels are more likely to adopt new technological solutions and innovations.

Li et al. (2020) concluded that one could see significant differences in attitudes to AI in the service encounter based on education. In a study where the AI chatbot's identity was initially concealed, the lower the educational level, the higher the likelihood of the conversation being immediately ceased following the revelation of the chatbot identity. Li et al. (2020) also concluded that this is a symptom of those with more education being solution-oriented, in contrast to those with less education who were less tolerant to the chatbot. That is, the higher the educational level, the more focused on the outcome of the service encounter consumers are.

Although these results are not specifically related to customer satisfaction and perceived employee performance, results have been shown and conclusions have been made within the service encounter that leads us to believe that our results will be similar. Based on this we theorize the following;

H4a: Education moderates customer satisfaction. In comparison with higher educated respondents, lower educated participants who have interacted with a human service employee will to a greater degree report higher customer satisfaction in contrast to those having interacted with a chatbot.

H4b: Education moderates customer satisfaction. In comparison with higher educated respondents, lower educated participants who have interacted with a human service employee will to a greater degree report higher perceived employee performance in contrast to those having interacted with a chatbot.

Note: we are aware that age, gender, and education put together could have interaction effects, but we choose not to examine any interactions effects between all the independent variables, just individually.

3. Methodology

3.1 Scientific approach and research strategy

The general approach to research in this area is quantitative studies, utilizing questionnaires. Söderlund's (2018) work is an example. Based on this and the following discussion, we decided that our ontological approach was to be a deductive one. Since the thesis is centered around a topic that is quite unexplored, with a technology that is not yet at a level where it can rival the interactions with another human on most platforms (Lafforgue, 2019), we leaned towards a deductive approach. In our reasoning we would be better able to understand the relatively unexplored topic if we conducted a quantitative study, since the variables we are measuring are almost always measured quantitatively.

We had already formed a perception of the area, as we had read several of Söderlund's works previously. Given this, we continued the same path and started with the theory. Following a positivistic point of view, as we knew what we wanted to study, and what we believed would be the outcomes of the study, we formulated the hypothesis.

We decided that the best way to study the phenomenons, and thus answer the hypotheses, was via a quantitative questionnaire. By using this method, we strove to capture an objective image of the attitudes that people have towards chatbots in customer service, and how that compares to the attitudes to chatting with a human service employee. This was a practical choice, as questionnaires are a simple, and fast, way of collecting data. Quantitative studies are also a good way of creating a perception of the general attitudes of a population. Further basing this choice on quantitative methodology, and given our ontological and epistemological assumptions, it was clear to us that this was a good method to follow through with the research project. Thus, the selected research strategy as well as scientific approach appeared to be natural choices. As we will return to, there are issues with these choices, especially considering the limitations of this thesis.

3.2 Main study

To explore the topic of algorithm aversion and whether there is an underlying difference in satisfaction and perceived employee performance for online customer service interactions with humans versus bots, we conducted a quantitative study through an experimental method, more specifically by conducting a survey through an online self-completion questionnaire (see Appendix 1 for the entire questionnaire). The reasoning behind using an experimental method was that we wanted to simulate a customer service interaction in an online setting and be able to compare the interactions with humans and with bots where the prerequisites are the same and a fair comparison can be made. In other words, a completely identical interaction.

3.2.1. Questionnaire

The study was conducted through a four-page self-completion questionnaire, based on a written case of an online customer service interaction, all written in English. On the first page, the respondent was introduced to the study. Information is provided on the aim of the questionnaire, the authors contact information, affirmation that the answers remain anonymous, and that participation is entirely voluntary. On the second page, the respondent encountered information regarding the handling of sensitive personal data, in line with GDPR regulations. This included information regarding the anonymity of the data, the secure storage of data, and the affirmation that no data published will be able to identify the respondent. To proceed beyond the second page, the respondent had to agree to the terms stated. On the third page, the respondents were presented with a written online customer service interaction that they are instructed to imagine themselves in, using a text-based role-play design. The fourth and fifth page of the survey consisted of ten questions, relating to the dependent variables measured.

3.2.2. Case scenario

The case scenario was constructed to simulate the experience of a successful and pleasant low complexity service interaction in an e-commerce setting through a written chat. The interaction is influenced by that used by Crolic et al. (2022). The customer, i.e., the respondent, reaches out to the customer service following a previous delivery system failure, a case in which the content of the employee response determines perceived customer satisfaction (Bitner et al., 1990). The reasoning behind providing a positive interaction was to reduce the likelihood of the respondent basing their answers on any previous negative experience with online customer service.

The respondent was asked to imagine themselves as the customer interacting with customer service. While the outcome and conversations were identical, respondents were randomly assigned to either a chatbot or a human employee using the randomizer option in Qualtrics. Thus, any differences in responses, customer satisfaction and perceived employee performance stems only from the fact that the messages are written by either a chatbot, or a human service employee. Thereby, any other differences that would naturally exist between the two, such as speed of response, accuracy, and empathy, have been eliminated. Furthermore, the fact that there were two different scenarios, employee or chatbot, was something the respondents were unaware of. To ensure that the participants were fully aware that they were interacting with an employee, or chatbot, the cases were designed so that the parties of the conversation were written in bold letters before each line of the conversation.

Further, to ensure that the scenario was relatable and realistic, the situation was designed to be something most people could see themselves in. The product ordered in the case was clothing for an upcoming wedding, something that most participants likely will be able to relate to regardless of culture or ethnicity as wedding ceremonies are something most are familiar with. The usage of a gender-neutral terminology in the item purchased in the case scenarios, “clothing” instead of for example dress or suit, was used to ensure that the scenario was realistic and applicable to people regardless of gender. Furthermore, a question asking whether the participant perceived the case to be realistic was added to the survey.

3.2.3 Questions

The survey contained questions to measure the variables customer satisfaction and perceived employee performance, with ten-point scales to answer the questions. To measure customer satisfaction, we used a variation of the three questions originally created by Fornell (1992): “How satisfied are you with the service interaction?” (1= very dissatisfied to 10= very satisfied), “How well did the service meet your expectations?” (1= not at all to 10= totally) and “How likely is it that you remember the interaction with the company as satisfactory?” (1= very unlikely to 10= very likely). As for measuring perceived performance the following of Söderlund’s (2018, p. 48) questions were used: “Please rate the customer service agent with respect to accessibility if you need help” (1= poor performance to 10= good performance), “Please rate the customer service agent with respect to helpfulness” (1= poor performance to 10= good performance) and “Please rate the customer service agent with respect to friendliness” (1= poor performance to 10= good performance). These questions were then followed by a question regarding whether the respondent interpreted the case scenario as realistic or not, which was also graded on a ten-point scale (1= very unrealistic to 10= very realistic).

The final section of the survey includes questions for three demographic variables: age, gender and educational level. The reasoning for including these demographic variables was that previous research has shown differences in attitudes to technology based on these three variables, as previously resonated.

The final question was a trap question (Jones et al., 2015), where the respondent had to select the number eight (8) from a choice of five numbers. If the wrong number was picked the response was deemed invalid and removed. This was to ensure that the respondents were not just typing in answers and thus increase the validity of the data.

3.3 Data collection and analysis

3.3.1 Data collection

The distribution of the survey started on the 15th of March 2022 and finished on the 8th of April 2022. A sum of 214 valid responses were recorded. The questionnaire was mainly distributed physically through the authors approaching respondents in public and providing the survey through a tablet device. Throughout all the collection of responses a single link has been used, and the randomization has then been made within Qualtrics, as described previously. The main space in which this took place was in the atrium of the Stockholm School of Economics, the SSE. As most of the answers were collected in this manner, this resulted in most respondents being students.

Furthermore, we took this way of collecting answers into account when we designed the questionnaire itself. This is reflected in the fact that we chose to have relatively few questions in the questionnaire. We believed that having fewer questions minimized the risk of those requested to answer the survey being stressed by the physical presence of either one of the authors, and thereby lowering the quality of the responses. Having fewer questions minimized the time spent and was meant to increase the quality of responses.

The starting point, and most of the answers we got initially, were from the students at the SSE. Since part of the thesis was to answer whether there were any differences in responses based on age, and educational level, this posed an issue. We realized that only collecting responses at the SSE did not provide a satisfactorily wide set of data for us to be able to see any differences other than based on gender. We concluded that we had to start collecting answers in other settings. Hence, we began publishing the survey in our private social media channels, such as Facebook and Linkedin, as well as sending it out directly to other colleagues, acquaintances, and family. In a lot of cases, we also asked them to further distribute the survey. By mainly focusing on sending it to those we know who are older than the student sample previously collected, this provided us with a slightly more extensive sample.

We continued by posting in various groups on Facebook (See Appendix 2 for full example of Facebook post). We utilized groups created with the goal of facilitating the collection of data through surveys and self-completion questionnaires. As we mainly needed older, highly educated, respondents, we published in groups where the focus was master's and PHD theses. In these groups, students, and professionals, "trade" answers with each other, by answering surveys and publishing their own with the hopes of others answering theirs in return. Through these groups on social media, we later found a webpage called Pollpool.com. The idea behind Pollpool is the same as in the Facebook groups, to enable the trade of survey answers. On this website you collect points by answering others' questionnaires. Points which we then could trade in for responses. Another advantage of this service was that we could sort our respondents on, for example, age. We thus chose to only receive answers from respondents aged 30 and older. This did succeed in providing us with enough answers in the other segments which we wished to study, however, as it would turn out, using these methods to reach the groups whom we could not approach at the SSE would bring other problems that we did not consider beforehand. We will return to these issues further.

The initial sample collected at the SSE had a high response rate, as no one declined to answer the survey. However, we began with a student convenience sample (Bell et al., 2019). As we collected more answers from different sources, the sample did grow significantly, but as we did not have to put in significant effort to reach these respondents, and since they are not to be considered representative of any population, these too are convenience samples.

3.3.2 Quality of data

In total, we collected 267 responses, out of which 214 were valid and used in the analysis. The total amount of responses included many which we were forced to remove for different reasons. Two answers were excluded as they had not finished the entire survey, by not answering all questions. 11 answers were excluded as they answered the control question incorrectly, thereby showing that they had not been paying attention to the questionnaire. 40 answers were excluded as they did not answer, or left incomprehensible answers, for example when asked what year they were born. For that question, we got faulty answers such as "Sweden". These answers showed that these respondents had not paid attention to the survey. Although they had answered the control question correctly, we excluded them. This question was in no way meant as a trick- or control question, however, it did function as one. The answers were divided 22 for chatbot, and 18 for the human employee. Keeping these respondents did not have a significant effect on the results.

As we have been present during the collection of answers at the SSE, we believe that most of the excluded answers come from the other means of collecting answers which have been used. Likely foremost from the answers collected via Pollpool, and the Facebook groups. This is one of the issues with providing a form of reward for each response. Another major issue being that it has created a bias in that group of respondents, which may have resulted in differences in the responses we have recorded. As we sorted on ages above 30 in Pollpool and aimed at older students when publishing in Facebook groups, these differences take place when comparing younger and older age groups. This may materialize in that those who have been awarded either feel generally more positive, or negative. However, what we believe is most likely is that the quality of these answers is worse. As these respondents can gather more answers to their own survey the more answers they provide themselves, they have an incentive to go through a survey as quickly as possible. We believe this is further reflected in the large number of faulty and excluded responses.

However, it is important to note that as we have randomized all respondents to the two cases, the sampling issues do not exist between the two experiment groups, and thus do not impact the results and conclusions of H1a and b. The potential harm is in the results and conclusions of H2a and b, as there may exist differences in the groups divided on age.

3.3.3 Data Analysis

The survey and gathering of responses were conducted using the online survey tool Qualtrics. The analysis of the responses was however done on IBM's SPSS Statistics tool, to where the data was exported. For starters, the data had to be cleaned, mainly by removing faulty answers. There were also three respondents who had misinterpreted the question regarding age, by not typing in year of birth and instead writing their age in years. We decided to correct the answers to the year of birth and keep the data as all other question responses were valid.

This was followed by summarizing the descriptive data, creating an overview of the respondents regarding what case they had encountered, the demographic data and the distribution of scenarios. We created multi-item scales for the three customer satisfaction and perceived employee performance questions respectively. To measure the internal consistency and scale reliability that is needed for the multi-item scales, we created a coefficient in the form of Cronbach's alpha. The result was a score of 0.84 for customer satisfaction, and a score of 0.9 for perceived employee performance, rounded to two decimal places, which suggests that the customer satisfaction and perceived employee performance items have a relatively high internal consistency, a score above 0.7 (Söderlund, 2005). After creating the multi-item scales, we ran independent *t*-tests, analyzing customer satisfaction and perceived employee performance in regards to the independent variables; demographics, and scenario, one by one.

To compare and analyze the independent variables of age and education, we decided to divide the data into two groups for age and two groups for education. Education was divided into "low education" (High school graduate + some college) and "high education" (college degree + doctorate). None of the other alternatives had any data. The division between participants who have not received and participants who have received a college degree was both a practical choice, as these groups became properly sized for analysis, but it also made sense since we had no further knowledge of how much, or little, education the "some college" group had. Age was divided into a "generation Z group" (people born 1995 and later) and an "old group" (people born 1994 and earlier). The decision to split the data into two groups for age was made since research points to a difference in attitude towards technology based on age. Generation Z are known as "digital natives" who are well accustomed to digital technology (Francis & Hoefel, 2018). As for the gender variable, we decided to remove the four respondents that had chosen not to state their gender, since the group was so small. However, we chose to include those four respondents in the other analyses.

This was followed by Univariate Two-way ANOVA analyzes to compare mean, standard deviation of satisfaction and perceived employee performance, how the independent variables affect the data and to see whether there were any significant differences or interaction effects stemming from the independent variables, one at a time, thus excluding to analyze for any interaction effects between the independent variables. We chose to conduct this type of test as it was most optimal for concluding whether the independent variables (age, gender, education and case) moderates satisfaction and perceived employee performance, thus being appropriate for our chosen hypotheses. The Two-way ANOVA analyses were done on the independent variables of the two scenarios, in combination with one of the demographic variables. The alternatives we considered were conducting two-way ANOVA with the independent variables as covariates, which would be appropriate if we were to test for covariance, or a multiple regression analysis, which is this case would be the more appropriate alternative. We chose not to conduct the multiple regression analysis as the added benefit of including all the variables that we hypothesized to be moderators and analyzing the interaction effects between all of them would not be relevant for the scope of our thesis.

3.4 Research reliability and validity

3.4.1. Reliability

Reliability is referred to when determining to what extent the same result of a study is received from several measurements (Bell et al., 2019), or the degree to which measures are free from error (Peter, 1979). The measurements used in this study have been utilized in several previous studies and have been proven to have a high degree of reliability.

We have used three questions to test customer satisfaction, and an additional three to test perceived employee performance, so called multi-item measures. They have been combined into multi-item scales and interpreted as a mean. The three measurements respectively have a Cronbach's Alpha-score above 0.7, where 0 is no internal reliability and 1 is perfect internal reliability (Bell et al., 2019). The consistency of the questions to each other is thus ensured, which is an indicator of general reliability (Söderlund, 2005). Lastly, type 1 errors are common in experiments. To reduce the negative effect this would have on reliability, a significance level of 5% was chosen.

3.4.2. Validity

External & internal validity

Validity refers to the degree to which the measure of a concept mirrors that concept or not (Bell et al., 2019).

Internal validity considers the causality of independent variables and whether they impact the variation in the dependent variable (Bell et al., 2019). To increase internal validity, the survey and case scenarios were assigned randomly to the respondents, with a randomizer option in Qualtrics. One pre-study was done to receive feedback on the case scenarios and the survey. The pre-study was conducted to make sure that the questions were easy to interpret, and that the respondents were aware that they interacted with a chatbot or employee. The pre-study revealed that respondents found the initial reading of the case scenario as something that made them want to skip parts of the survey, as it was perceived as quite long, thus prompting us to include a control question. This control question needed to be answered correctly to prove that the participant was paying attention to the study. Faulty answers were removed from the study, increasing the internal validity.

External validity refers to the participants in the study and whether the findings from this sample can be generalized beyond the area of the study (Bell et al., 2019). Because of the use of a convenience sample, where many of the participants have similar background and demographic characteristics, and participants were chosen based on accessibility for the most part, the participants in the study are not representative of any population. As a result, the study's external validity is limited. Furthermore, as the case used was very specific, and since it was written to convey a positive experience, generalizability to other contexts, and experiences, within customer service is further decreased. Lastly, as the study only looked at two aspects of attitude, customer satisfaction and perceived employee performance, the results are not generalizable to other measurements of attitude, or attitude as a broader concept.

Replicability

The measurements and scales used in this study have been used in previous papers, on which this study builds. As roughly half of the participants “interacted” with a bot, we had to adapt the questions to also fit that scenario. We thus used the terminology “customer service agent” for the questions regarding perceived employee performance, instead of “personnel”, which was used by Söderlund (2018). Furthermore, we chose to only use three out of Söderlund’s (2018) six questions, as we believed they were most relevant. The other three of Söderlund’s (2018, p. 48) questions that we chose to exclude are “Please rate the personnel with respect to…”, with “attention”, “interest for you as a customer” and “knowledge if you ask about something”, which we deemed had no practical fit to the case scenarios.

3.4.3. Perceived realisticity

To increase understanding of how respondents reacted to the cases, we included a question at the end of the survey to determine the level of perceived realisticity of the scenario. The question “How realistic do you think the interaction was?”, which was answered on a 10-point scale (1= Very unrealistic to 10= Very realistic), does not necessarily have a high degree of reliability, as we did not use a multi-item scale, but did provide some indication of validity. Running a *t*-test on the question of how realistic the respondent perceived the interaction to be, we found that both cases scored very similar to each other. The data presented in table 1 showed no significant difference in perceived realisticity when comparing the two cases independent of any other variable. ($t(212) = 1.759$, $p = 0.935$).

The respondents interacting with an employee did on average rate the realism 5.98/10, while chatbot respondents on average scored 6.01/10, indicating that both scenarios were perceived as neither realistic nor unrealistic.

Table 1: Independent T-test on perceived realisticity chatbot versus employee

Scenario	Total N=214	Mean	Std. Deviation	Significance Two-sided
Employee	109	5.98	2.38	0.935
Chatbot	105	6.01	2.61	

Note: The scale for the question was from 1= very unrealistic, to 10= very realistic.

4. Results

4.1 Descriptive statistics

The total number of valid responses was 214. 107 of the respondents were female (50%), while 102 (47.7%) respondents were male. No responses were recorded for the "other" option, whilst five respondents (2.3%) preferred not to state their gender. As for education, the largest recorded group (121 respondents), had obtained a college degree (56.5%). "Some college" was the second largest group with 59 respondents (27.6%), "High school" followed with 27 respondents (12.6%) and lastly, "Doctorate" with 7 respondents (3.3%). The age distribution shows that most responses were made by people aged 18-25 with 123 answers (57.5%). This is followed by people aged 26-35 with 39 respondents (18.2%), aged 56-65 with 30 respondents (14%), aged 36-45 with 12 respondents (5.6%), aged 65 and older with 7 respondents (3.3%), and lastly aged 46-55 with 3 respondents (1.4%). Thus, for the group labeled "Generation Z", there were 139 respondents, and 75 for the older group

Table 2: Gender, education, and age distribution

Gender	Total N=214	% of total sample	Employee	Chatbot
Male	102	48		
Female	107	50		
Other	0	0		
Prefer not to say	5	2		
Education	**Total N=214**	**% of total sample**		
<High school	0	0		
High school	27	13		
Some college	59	28		
College	121	57		
Doctorate	7	3		
Age (years)	**Total N=214**	**% of total sample**		
18-25	123	58		
26-35	39	18		
36-45	12	6		
46-55	3	1		
56-65	30	14		
>65	7	3		
Group	**Total N=214**	**% of total sample**	**Employee**	**Chatbot**
Male	102	48	55	47
Female	107	50	51	56
Less educated	86	40	43	43
Educated	128	60	66	62
Gen Z	139	65	76	63
Older	75	35	33	42

Note: The percentages were rounded towards the nearest integer.

4.2 Satisfaction and perceived employee performance

In H1a and H1b we predicted that customer satisfaction and perceived employee performance would be higher for the group that had interacted with a human employee, compared to the group interacting with a chatbot. To analyze the differences in customer satisfaction and perceived employee performance between the two case groups, an independent *t*-test was conducted.

As for customer satisfaction, the data presented in table 3 showed no significant difference in customer satisfaction when comparing the two cases independent of any other variable ($t(213) = 1.356$, $p = 0.088$).

This result was also the same for perceived employee performance presented in table 3, where no significant difference between the groups were found ($t(213) = 1.1$, $p = 0.136$). We could thus not find support for H1a and H1b.

Table 3: Independent T-test on satisfaction and performance index, chatbot versus employee

Scenario	Employee M (SD)	Chatbot M (SD)	T-value	Significance
Satisfaction	8.73 (1.53)	8.43 (1.66)	1.356	0.088
Performance	8.86 (1.36)	8.64 (1.54)	1.1	0.136

4.3 Age

In H2a and H2b we predicted that customer satisfaction and perceived employee performance would be higher for the Gen Z group that had interacted with a human employee, compared to a chatbot, with age acting as a moderator for customer satisfaction and perceived employee performance. A two-way ANOVA which examined the effects of the different cases (chatbot and employee) and age on customer satisfaction was conducted. The data presented in table 4 showed no significant differences in customer satisfaction. This result was also the same for perceived employee performance presented in table 5, where no significant differences between the groups was found. Neither were there significant interaction effects for the independent variable of age. We could not find support for H2a and H2b. Thus, we cannot conclude that age has a moderating effect on either customer satisfaction or perceived employee performance.

As for customer satisfaction, the overall results of the two-way ANOVA were not significant ($F(3, 213) = 1.259$, $p = 0.289$). In other words, there was no significant difference in customer satisfaction between the two cases. There was no significant main effect of age ($F(1, 213) = 1.774$, $p = 0.184$) nor the two cases ($F(1, 213) = 1.585$, $p = 289$). No significant interaction effect between the two variables could be found ($F(1,213) = 0.120$, $p = 0.730$). Thus, no statistically significant evidence could be found for H2a, meaning that we could not find support for age acting as a moderator for satisfaction.

Table 4: Univariate ANOVA test on customer satisfaction and age

Group	Satisfaction employee			Satisfaction chatbot		
	Mean	**SD**	**(n)**	**Mean**	**SD**	**(n)**
Old group	8.55	1.69	(33)	8.11	2.03	(42)
Gen Z	8.8	1.4	(76)	8.63	1.38	(63)

Note: The main effect of the case scenarios on satisfaction was not significant (p = .209), neither was the main effect of the age group (p = .184) or the interaction between age group and the case scenario (p = .730).

The results of the two-way ANOVA for perceived employee performance were not significant (F(3, 213) = 1.181, p = 0.318). No significant main effect of age could be found (F(1, 213) = 1.945, p = 0.165), nor could any significant effect of the two cases be found (F(1, 213) = 1.154, p = 0.284). No significant interaction effect between the two variables could be found (F(1, 213) = 0.310, p = 0.578). The two-way ANOVA analysis has thus shown that no statistically significant evidence could be found for H2b, there was no evidence found that age moderates perceived employee performance.

Table 5: Univariate ANOVA test on perceived employee performance and age

Group	Performance employee			Performance chatbot		
	Mean	**SD**	**(n)**	**Mean**	**SD**	**(n)**
Old group	8.76	1.56	(33)	8.31	1.83	(42)
Gen Z	8.89	1.29	(76)	8.84	1.31	(63)

Note: The main effect of the case scenarios on performance was not significant (p = .284), neither was the main effect of the age group (p = .165) or the interaction between age group and the case scenario (p = .578).

4.4 Gender

In H3a and H3b we predicted that gender moderates customer satisfaction and perceived employee performance. To analyze the effects of the two cases (chatbot and employee) and gender (male and female) on customer satisfaction and perceived employee performance, a two-way ANOVA was performed.

For customer satisfaction, the results of the two-way ANOVA were not significant (F(3, 213) = 1.299, p = 0.276). There was no significant main effect of gender (F(1, 213) = 0.148, p = 0.701) nor the two cases (F(1, 213) = 1.926, p = 167). No significant interaction effect between the two variables could be found (F(1, 213) = 1.820, p = 0.179). Thus, no statistically significant support could be found for H3a, meaning that we could not find support for gender having a moderating effect on customer satisfaction.

Table 6: Univariate two-way ANOVA test on customer satisfaction and gender

Group	Satisfaction employee			Satisfaction chatbot		
	Mean	SD	(n)	Mean	SD	(n)
Female group	8.64	1.83	(51)	8.14	2.06	(56)
Male group	8.80	1.40	(55)	8.63	1.38	(47)

Note: The main effect of the case scenarios on satisfaction was not significant (p = .167), neither was the main effect of gender (p = .701) or the interaction between gender and the case scenario (p = .179).

Neither were the results of the two-way ANOVA for perceived employee performance significant (F(3, 213) = 0.570, p = 0.635). No significant main effect of gender could be found (F(1, 213) = 0.019, p = 0.890), nor could any significant effect of the two cases be found (F(1, 213) = 1.661, p = 0.199). No significant interaction effect between the two variables could be found either (F(1, 213) = 0.001, p = 0.981). The two-way ANOVA analysis has thus shown that no statistically significant support could be found for H3b. There was no support that gender moderates perceived employee performance.

Table 7: Univariate ANOVA test on perceived employee performance and gender

Group	Performance employee			Performance chatbot		
	Mean	SD	(n)	Mean	SD	(n)
Female group	8.89	1.16	(51)	8.66	1.22	(56)
Male group	8.89	1.46	(55)	8.64	1.78	(47)

Note: The main effect of the case scenarios on perceived employee performance was not significant (p = .199), neither was the main effect gender (p = .890) or the interaction between gender and the case scenarios (p = .981).

4.5 Education

In H4a and H4b we predicted that education moderates customer satisfaction and perceived employee performance, respectively. To analyze the effects of the two cases (chatbot and employee) and education (low and high) on customer satisfaction and perceived employee performance, a two-way ANOVA was performed.

As for customer satisfaction, the overall results of the two-way ANOVA were not significant (F(3, 213) = 0.875, p = 0.439). There was no significant main effect of education (F(1, 213) = 0.488, p = 0.385) nor the two cases (F(1, 213) = 1.501, p = 186). No significant interaction effect between the two variables could be found (F(1, 213) = 0.320, p = 0.796). Thus, no statistically significant support could be found for H4a, we could not find support for education acting as a moderator of customer satisfaction.

Table 8: Univariate ANOVA test on customer satisfaction and education

Group	Satisfaction employee			Satisfaction chatbot		
	Mean	**SD**	**(n)**	**Mean**	**SD**	**(n)**
Low education	8.75	1.49	(43)	8.60	1.45	(43)
High education	8.72	1.57	(66)	8.32	1.8	(62)

Note: The main effect of the case scenarios on satisfaction was not significant (p = .186), neither was the main effect of the education group (p = .385) or the interaction between education and the case scenario (p = .796).

The results of the two-way ANOVA for perceived employee performance were not significant ($F(3, 213) = 1.045$, $p = 0.287$). No significant main effect of education could be found ($F(1, 213) = 0.797$, $p = 0.199$), nor could any significant effect of the two cases be found ($F(1, 213) = 0.777$, $p = 0.235$). No significant interaction effect between the two variables could be found ($F(1, 213) = 1.152$, $p = 0.512$). The two-way ANOVA analysis has thus shown that there was no statistically significant support for H4b, there was no support of education moderating perceived employee performance.

Table 9: Univariate ANOVA test on perceived employee performance and education

Group	Performance employee			Performance chatbot		
	Mean	**SD**	**(n)**	**Mean**	**SD**	**(n)**
Low education	8.84	1.35	(43)	8.88	1.27	(43)
High education	8.87	1.38	(66)	8.48	1.69	(62)

Note: The main effect of the case scenarios on perceived employee performance was not significant (p = 0.235), neither was the main effect of the education group (p = 0.199) or the interaction between education and the case scenario (p = 0.512).

5. Discussion

The purpose of the thesis is to examine how consumers perceive service encounters from chatbots and employees. In particular, we examine whether consumers will rate customer satisfaction and perceived employee performance differently when having been in contact with a chatbot and a human. We further strive to conclude what potential differences, if any, there are in the responses of different demographic groups, divided on age, education, and gender. Based on this purpose the research question we aim to answer is therefore the following:

To what extent do consumers rate human-to-human-interaction in e-commerce customer encounters as more positively than interacting with a chatbot? What factors explain this potential difference? To what extent could demographic variables moderate this difference?

5.1 Summary of results

Table 10: Summary of hypotheses and conclusions from results

Hypothesis	Result
H1a	Not supported
H1b	Not supported
H2a	Not supported
H2b	Not supported
H3a	Not supported
H3b	Not supported
H4a	Not supported
H4b	Not supported

The concluding findings from the data analysis using independent *t*-tests and two-way ANOVA show that no significant support for any of the study's hypotheses could be found. Thus, there was no significant support for any difference in mean for customer satisfaction and perceived employee performance between the two cases (chatbot and employee). In other words, there was no significant difference in how people rated the different case scenarios, chatbot and employee, on customer satisfaction and perceived employee performance. Further, the results showed no statistically significant support for either age, education nor gender acting as mediators for customer satisfaction and perceived employee performance.

5.2 Conclusion and implications

The general conclusion to be drawn is that, based on our sample and in the specific case of this study, there are no significant differences between the degree of customer satisfaction and perceived employee performance after having interacted with either a chatbot, or a human service employee. This study was limited in determining the factors behind the results that we got. However, based on the theory on which it was based, there are some conclusions that can be drawn, and possible explanations to our results. However, as the degree of external validity, and generalizability, as has been discussed before, is arguably quite low, the generalizations made here are foremost applicable in this specific type of service encounters.

As both scenarios were the same apart from who responded on the other end, the degree of unsystematic variability, as described by Söderlund (2012), was low. While service heterogeneity generally refers to repeated interactions, our results seem to be in line with the conclusion that low degrees of unsystematic variability results in high degrees of customer satisfaction. Furthermore, the level of control for the respondent was low, which should result in negative attitudes (Söderlund, 2012). Yet, as Ridgeway (1987) claimed, our results imply that consumers are willing to give up control to allow the task to be solved.

Algorithm aversion, that the choice between a human and an algorithm is not based on rational, objective criteria (Jussupow et al., 2020), is not present in this instance. Consumer do not prefer humans. This may be a sign that intermediary factors, such as trust, are valued on par between the bot and the human. Furthermore, we can conclude that in this case consumers do not rely on their previous beliefs about chatbots, as claimed by Mozafari et al. (2021). Alternatively, previous perceptions of bots are on the same level as those of human employees. It is reasonable to conclude that people have higher degrees of trust in bots than before. Users expect bots to be as well trained, and the outcome of interacting with a bot to be as predictable and stable as when interacting with a human. The development and quality of AI is moving quickly, and it is rational to assume that the expectations on them are moving as quickly, along the lines of the conclusions made by Collier & Kimes (2013) and Grewal et al. (2017).

As with the claims of algorithm aversion, we neither found evidence of algorithm appreciation being present in our sample set. If anything, our results contribute to the conclusion that these factors play different roles in different situations. As previously established, the task in the case was of low complexity, which Xu et al. (2020) claimed results in higher intent of usage of bots. To some degree our results coincide with this. For low complexity tasks, consumers are as satisfied with bots as they are with humans.

That consumers do not rate chatbots any less positive than human employees provide many implications for e-commerce. As already discussed, there is a rift between practice and theory within this area. Companies are heavily investing in, and employing, chatbots (Belanche et al., 2020; Lu et al., 2020), while theory on algorithms claim that customers rate them worse than humans. It is within the nature of business to cut costs wherever possible, to increase profits. By employing chatbots, companies can employ less people, which results in less wages being paid out (De, 2018). Chatbots are also more stable in comparison to humans, which results in lower levels of unsystematic variability connected to higher customer satisfaction (Söderlund, 2012). In combination, this substantiates the claims made by De (2018), that chatbots will improve the quality of service while reducing costs.

This area is developing extremely fast, and so is the general perception, as well as consumers' attitudes to it (Collier & Kimes, 2013; Grewal et al., 2017). Theory has not managed to keep up with this development, and algorithm aversion may no longer be prevalent in the population of consumers. The companies might have already understood this, and that they should focus on the quality of the service, and solving issues, rather than on who does the solving. Consumers seem to be more concerned with the outcome rather than the means of getting there. As there is no bias towards chatbots, and they provide significant cost savings, it is beneficial to utilize them.

5.2.1. Conclusions and implications on age, gender, and education

The digital divide between generations, older generations, , defined by Barbosa Neves et al. (2018) as 65 and older, being more technophobic and less willing to adopt digital communication tools (Barbosa Neves et al., 2018; Berkowsky et al., 2015), is not prevalent in our sample. This may be explained by the fact that older generations do not experience any difference between conversing through chat with a human or bot. One might have expected then that older generations' attitudes would have been less positive in general. The results of non-significant independent *t*-tests, however, showed that there were no such significant differences in customer satisfaction nor perceived employee performance. There may be no differences since the scenario was not an actual chat, indicating that older generations feel as comfortable in reading a conversation as younger generations do. The results might have been different had we been able to create an actual chat, which the respondents had to actively participate in.

Barbosa Neves et al. (2018) utilized a case study approach with an Australian and Canadian sample, with interviews and field observations. This differs from our study in not only the definition of older and younger, but also in sampling and general approach of the study.

Neither did we find any evidence of a gender digital divide. This underlines the claims that this gap has narrowed significantly and may indeed be an indication of the more equal degree of education and employment opportunities within the area of complex technology (Anthony Giddens, 2013; Kotze et al., 2016), substantiated by there being no significance in the independent *t*-tests done on gender. In this aspect, there seems to be less inequalities than historically, in line with more recent literature. Based on the structuration theory (Giddens, 2013; Giddens, 1984), we can conclude that social structures do not prohibit, or limit, women, in this narrow aspect and case.

The division based on education was arguably limiting. Had we got more answers, the division, and results, could have been more nuanced. However, we did not find any evidence of education acting as a mediator to customer satisfaction nor perceived employee performance. Neither did the independent *t*-tests on the two groups show anything with significance. This may, again, be a result of there not being a big difference between those with some college-, and a finished college education.

5.3 Key results and implications

In summary, the results of this study showed no differences in either customer satisfaction nor perceived employee performance whether respondents had read a conversation with a customer service employee, or chatbot. Neither did the hypothesized moderating factors show any differences in results.

The practical implication of this is to shift as many service encounters as possible into using AI. All chat interactions, at least those of low and medium task complexity, should be conducted by chatbots (Mozafari et al., 2021; Xu et al., 2020). This will increase the service quality, and provide cost saving opportunities (De, 2018). The money saved on less employees can instead be spent on other means of delivering satisfactory outcomes, beyond the area of this study.

5.4 Limitations and suggested improvements

As there were several limitations with the conduction of this study, so are there several possible improvements, for the recreation of this study, or future research. In hindsight, the questionnaire of the study could have been complemented with a manipulation check, designed to make sure the respondent interpreted the written case correctly. More precisely if they understood that they had interacted with a chatbot or employee, as this was central to the responses gathered in the study. Further, questions regarding the respondent’s experience with e-commerce, online customer service with humans and particularly with chatbots would have been a great addition to the questionnaire. Since these types of questions weren’t present, we could not gain insight into whether the respondent understands for example what a chatbot is.

As for improvements, the case could have been more interactive, giving a more realistic, actual, conversation. This would have given a more accurate and possibly more nuanced result. More answers should also have been gathered, especially among the older, and higher educated, groups. This would have enabled a finer division to analyze and would also not have led us into using Pollpool, for example, which resulted in the issue of different groups of respondents being treated differently.

There are also several other measures that should have been tested to get a better understanding of the motivation behind the responses received. As the employee is the service from the customer’s point of view, the effects of which are transferred to the company according to the affect infusion model (Ward & Barnes, 2001), testing brand image would have enabled the determination of whether the same level of customer satisfaction between groups is reflected when viewing the company. Furthermore, trust, the experienced authenticity of actions, experienced social distance, as well as perceived knowledge and empathy, some of the factors said to be intermediary to customer satisfaction (Jago, 2019; Luo et al., 2019; Madhavan & Wiegmann, 2007; Trope & Liberman, 2010; Önkal et al., 2009), would have been interesting to test to conclude further on the causes of our results. As we did firmly expect that we would get results confirming our hypothesis, we chose not to extend the study with further questions, which in hindsight would have made the discussion more interesting.

Further, something that could have added more nuance to the study is to have included an unpleasant case scenario, testing the exact same variables but for a negative interaction. Since not every service interaction ends up pleasant, the perspective of a negative outcome case scenario would surely have added nuance to the study.

Lastly, what would perhaps have contributed more than the previously mentioned additional metrics is to have tested expectations and previous attitudes before the interaction, to be able to make more elaborate conclusions about previous beliefs, whether and how it differed as well as whether it made any differences on the results. By doing this both through measurements, and open questions, we perhaps would have been able to contribute with more extensive knowledge than was now possible.

6. References

Anderson, J. C., & Narus, J. A. (1990). A model of distributor firm and manufacturer firm working partnerships. *Journal of Marketing, 54*(1), 42. 10.2307/1252172

Anthony Giddens, G. (2013). *The Constitution of Society: Outline of the Theory of Structuration*. Polity Press.

Barbosa Neves, B., Waycott, J., & Malta, S. (2018). Old and afraid of new communication technologies?: Reconceptualising and contesting the 'age-based digital divide'. *Journal of Sociology (Melbourne, Vic.), 54*(2), 236-248. 10.1177/1440783318766119

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success. *Journal of Service Management, 31*(2), 267-289. 10.1108/JOSM-05-2019-0156

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (Fifth edition ed.). Oxford University Press.

Berkowsky, R. W., Rikard, R. V., & Cotten, S. R. (2015). Signing Off: Predicting Discontinued ICT Usage Among Older Adults in Assisted and Independent Living. *Human Aspects of IT for the Aged Population. Design for Everyday Life* (pp. 389-398). Springer International Publishing. 10.1007/978-3-319-20913-5_36

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21-34. 10.1016/j.cognition.2018.08.003

Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: diagnosing favorable and unfavorable incidents. *Journal of Marketing, 54*(1), 71. 10.2307/1252174

Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: disrupting what we know about services. *The Journal of Services Marketing, 34*(3), 317-334. 10.1108/JSM-01-2019-0047

Broadbent, E. (2017). Interactions With Robots: The Truths We Reveal About Ourselves. *Annual Review of Psychology, 68*(1), 627-652. 10.1146/annurev-psych-010416-043958

Cameron, D. D., Marquis, D. R., & Webster, B. (2001). Older adults perceptions, experiences and anxieties with emerging technologies. *Australasian Journal on Ageing, 20*(s2), 50-56. 10.1111/j.1741-6612.2001.tb00399.x

Castelo, N., Bos, M. W., & Lehmann, D. (2019). Let the Machine Decide: When Consumers Trust or Distrust Algorithms. *NIM Marketing Intelligence Review, 11*(2), 24-29. 10.2478/nimmir-2019-0012

Chevalier, S. (2022). *Retail e-commerce sales worldwide from 2014 to 2025.* Statista. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

Christy, K. R., Jensen, J. D., Britt, B., Scherr, C. L., Jones, C., & Brown, N. R. (2019). I want to talk to a real person: theorising avoidance in the acceptance and use of automated technologies. *Health Systems, 8*(1), 31-43. 10.1080/20476965.2017.1406568

Cialdini, R. B. (2007). *Influence: the psychology of persuasion*. Collins.

Collier, J. E., & Kimes, S. E. (2013). Only if it is convenient. *Journal of Service Research, 16*(1), 39-51. http://www.econis.eu/PPNSET?PPN=749356022

Cramer-Flood, E. (2022). *Global Ecommerce Forecast 2022* . eMarketer. https://www.emarketer.com/content/global-ecommerce-forecast-2022

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions. *Journal of Marketing, 86*(1), 132-148. 10.1177/00222429211045687

De, A. (2018). *A Look at the Future of Chatbots in Customer Service.* ReadWrite [BLOG]. https://search.proquest.com/docview/2252760761

Deloitte, D. (2018). *Chatbots point of view.* Deloitte. https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/deloitte analytics/deloitte-nl-chatbots-moving-beyond-the-hype.pdf

Epley, N. (2018). A Mind like Mine: The Exceptionally Ordinary Underpinnings of Anthropomorphism. *Journal of the Association for Consumer Research, 3*(4), 591-598. 10.1086/699516

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*(1), 39-66. 10.1037//0033-2909.117.1.39

Fornell, C. (1992). A National Customer Satisfaction Barometer: The Swedish Experience. *Journal of Marketing, 56*(1), 6-21. 10.1177/002224299205600103

Foxall, G. R., & Greenley, G. E. (1999). Consumers' Emotional Responses to Service Environments. *Journal of Business Research, 46*(2), 149-158. 10.1016/S0148-2963(98)00018-6

Francis, T., & Hoefel, F. (2018). *The influence of Gen Z—the first generation of true digital natives—is expanding.* Mckinsey.com. https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/true-gen-generation-z-and-its-implications-for-companies

Giddens, A. (1984). *The constitution of society* (1. publ ed.). Wiley.

Gilbert, D., Lee-Kelley, L., & Barton, M. (2003). Technophobia, gender influences and consumer decision-making for technology-related products. *European Journal of Innovation Management, 6*(4), 253-263. 10.1108/14601060310500968

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing, 93*(1), 1-6. 10.1016/j.jretai.2016.12.008

Hadi, R. (2019). When Humanizing Customer Service Chatbots Might Backfire. *NIM Marketing Intelligence Review, 11*(2), 30-35. 10.2478/nimmir-2019-0013

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Hoffman, K. D., & Bateson, J. E. G. (1997). *Essentials of services marketing*. Dryden Press.

Huang, M., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research : JSR, 21*(2), 155-172. 10.1177/1094670517752459

Hudson, J., Orviska, M., & Hunady, J. (2019). People's attitudes to autonomous vehicles. *Transportation Research. Part A, Policy and Practice, 121*, 164-176. 10.1016/j.tra.2018.08.018

Insider, I. (2022). *Chatbot market in 2022: Stats, trends, and companies in the growing AI chatbot industry.* Business Insider. https://www.businessinsider.com/chatbot-market-stats-trends?r=AU&IR=T

Jago, A. S. (2019). Algorithms and Authenticity. *Academy of Management Discoveries, 5*(1), 38-56. 10.5465/amd.2017.0002

Jin-Jong, C. (2015). Overcoming technophobia in poorly-educated elderly – the HELPS-seniors service learning program. *International Journal of Automation and Smart Technology, 5*(3), 173-182. 10.5875/ausmt.v5i3.980

Jones, T. O. (1996). Why Satisfied Customers Defect. *Journal of Management in Engineering, 12*(6), 11. 10.1061/(ASCE)0742-597X(1996)12:6(11.2)

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion. Paper presented at the https://madoc.bib.uni-mannheim.de/56152/

Kotler, P. (1994). Reconceptualizing marketing: An interview with Philip Kotler. *European Management Journal, 12*(4), 353-361. 10.1016/0263-2373(94)90021-3

Kotze, T. G., Anderson, O., & Summerfield, K. (2016). Technophobia : gender differences in the adoption of high-technology consumer products. *South African Journal of Business Management, 47*(1), 21-28. 10.4102/sajbm.v47i1.49

Lafforgue, B. (2019). *CHATBOT LEVELS : HOW TO CHOOSE YOUR AI VIRTUAL AGENT ?* Konverso.ai. https://www.konverso.ai/blog/chatbot-level-choose-ai-virtual-agent

Li, C., Pan, R., Xin, H., & Deng, Z. (2020). Research on Artificial Intelligence Customer Service on Consumer Attitude and Its Impact during Online Shopping. *Journal of Physics. Conference Series, 1575*(1), 12192. 10.1088/1742-6596/1575/1/012192

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90-103. 10.1016/j.obhdp.2018.12.005

Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? *Journal of Service Theory and Practice, 30*(3), 361-391. 10.1108/JSTP-04-2019-0088

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science (Providence, R.I.), 38*(6), 937-947. 10.1287/mksc.2019.1192

Madhavan, P., & Wiegmann, D. A. (2007). Effects of Information Source, Pedigree, and Reliability on Operator Interaction With Decision Support Systems. *Human Factors, 49*(5), 773-785. 10.1518/001872007X230154

Marescotti, M. E., Demartini, E., Filippini, R., & Gaviglio, A. (2021). Smart farming in mountain areas: Investigating livestock farmers' technophobia and technophilia and their perception of innovation. *Journal of Rural Studies, 86*, 463-472. 10.1016/j.jrurstud.2021.07.015

Marshall, K. (2017). *AI to power majority of customer interactions by 2025.* It Pro. https://www.itpro.co.uk/strategy/28281/ai-to-power-majority-of-customer-interactions-by-2025

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses. *Journal of Marketing Research, 56*(4), 535-556. 10.1177/0022243718822827

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management, ahead-of-print*(ahead-of-print)10.1108/JOSM-10-2020-0380

Murphy, J., Gretzel, U., & Pesonen, J. (2019). Marketing robot services in hospitality and tourism: the role of anthropomorphism. *Journal of Travel & Tourism Marketing, 36*(7), 784-795. 10.1080/10548408.2019.1571983

Neves, B. B., Franz, R. L., Munteanu, C., & Baecker, R. (2018). Adoption and feasibility of a communication app to enhance social connectedness amongst frail institutionalized oldest old: an embedded case study. *Information, Communication & Society, 21*(11), 1681-1699. 10.1080/1369118X.2017.1348534

Nimrod, G. (2018). Technophobia among older Internet users. *Educational Gerontology, 44*(2-3), 148-162. 10.1080/03601277.2018.1428145

Nix, E. (2018). *The World's First Web Site.* History. https://www.history.com/news/the-worlds-first-web-site

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making, 22*(4), 390-409. 10.1002/bdm.637

Oracle, O. (2022). *What is a chatbot?* Oracle. https://www.oracle.com/in/chatbots/what-is-a-chatbot/

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2018). Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences. *Handbook of Service Science, Volume II* (pp. 77-103). Springer International Publishing. 10.1007/978-3-319-98512-1_5

Palmeira, M., & Spassova, G. (2015). Consumer reactions to professionals who use decision aids. *European Journal of Marketing, 49*(3/4), 302-326. 10.1108/EJM-07-2013-0390

Peter, J. P. (1979). Reliability: A Review of Psychometric Basics and Recent Marketing Practices. *Journal of Marketing Research, 16*(1), 6-17. 10.1177/002224377901600102

Phaneuf, A. (2022). *Ecommerce Statistics: Industry benchmarks & growth* . Insider Intelligence. https://www.insiderintelligence.com/insights/ecommerce-industry-statistics/

Rafaeli, A., & Pratt, M. G. (1993). Tailored Meanings: On the Meaning and Impact of Organizational Dress. *The Academy of Management Review, 18*(1), 32-55. 10.2307/258822

Ridgeway, C. L. (1987). Nonverbal Behavior, Dominance, and the Basis of Status in Task Groups. *American Sociological Review, 52*(5), 683-694. 10.2307/2095603

Salesforce, S. (2021). *E-handel i Sverige 2021: statistik, historisk utveckling & framtidsperspektiv.* Salesforce. https://www.salesforce.com/se/blog/2021/09/e-handel-i-sverige-2021-statistik-historisk-utveckling-och-framtidsperspektiv.html

Scwab, K. (2016). *The Fourth Industrial Revolution: what it means, how to respond.* World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Sears, D. O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology, 44*(2), 233-250. 10.1037/0022-3514.44.2.233

Sheikh, J. A., & Abbas, A. (2015). The Invisible User: Women in DUXU. *Design, User Experience, and Usability: Users and Interactions* (pp. 243-251). Springer International Publishing. 10.1007/978-3-319-20898-5_24

Sumpter, D.J.T., (2018). *Outnumbered.* London ; Oxford ; New York ; New Delhi ; Sydney: Bloomsbury Sigma.

Söderlund, M. (2005). *Mätningar och mått* (1. uppl. ed.). Liber ekonomi.

Söderlund, M. (2012). *Kundmötet* (1. uppl. ed.). Liber.

Söderlund, M. (2018). The proactive employee on the floor of the store and the impact on customer satisfaction. *Journal of Retailing and Consumer Services, 43,* 46-53. http://www.econis.eu/PPNSET?PPN=1028405693

Söderlund, M. (2021). The robot-to-robot service encounter: an examination of the impact of inter-robot warmth. *The Journal of Services Marketing, 35*(9), 15-27. 10.1108/JSM-01-2021-0006

Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management, 69,* 135-146. 10.1016/j.indmarman.2017.12.019

Tran, A. D., Pallant, J. I., & Johnson, L. W. (2021). Exploring the impact of chatbots on consumer sentiment and expectations in retail. *Journal of Retailing and Consumer Services, 63,* 102718. 10.1016/j.jretconser.2021.102718

Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance. *Psychological Review, 117*(2), 440-463. 10.1037/a0018963

Vines, J., Pritchard, G., Wright, P., Olivier, P., & Brittain, K. (2015). An Age-Old Problem. *ACM Transactions on Computer-Human Interaction, 22*(1), 1-27. 10.1145/2696867

Ward, J. C., & Barnes, J. W. (2001). Control and affect: the influence of feeling in control of the retail environment on affect, involvement, attitude, and behavior. *Journal of Business Research, 54*(2), 139-144. 10.1016/S0148-2963(99)00083-1

Xu, Y., Shieh, C., van Esch, P., & Ling, I. (2020). AI customer service: Task complexity, problem-solving ability, and usage intention. *Australasian Marketing Journal, 28*(4), 189-199. 10.1016/j.ausmj.2020.03.005

Zwass, V. (2019). *e-commerce.* Encyclopedia Britannica. https://www.britannica.com/technology/e-commerce

7. Appendix

7.1 Appendix 1

How do you experience this customer service meeting?

Welcome to our study!

The study aims to gather knowledge about how people think about and experience meetings with customer service in e-commerce. It takes about 5 minutes to read the text and answer the web survey. Even an uncertain answer is of interest and better than no answer at all. Some questions may remind you of others.

All answers will be treated confidentially and will of course remain anonymous. The survey is part of a research project from the Stockholm School of Economics. The results of the project will be presented in a bachelors thesis.

If you have questions about the web survey or study, please contact us at the e-mail address: 24876@student.hhs.se

Many thanks in advance for your participation,

Arvid Bergstrand, student bachelor's program in Business & Economics, Stockholm School of Economics

Youcef Djehiche, student bachelor's program in Business & Economics, Stockholm School of Economics

Info relating to GDPR:

As an integral part of educational program at the Stockholm School of Economics, enrolled students complete a thesis. This work is sometimes based upon surveys. Participation is naturally entirely voluntary, and this text is intended to provide you with necessary information about that may concern your participation in the study. You can at any time withdraw your consent and your data will thereafter be permanently erased.

Confidentiality: Anything you say or state in the survey or to the interviewers will be held strictly confidential and will only be made available to supervisors, tutors and the course management team.

Secure storage of data: All data will be stored and processed safely by the SSE and will be permanently deleted when the project is completed.

No personal data will be published: The thesis written by the students will not contain any information that may identify you as participant to the survey or subject.

Your rights under GDPR: You are welcome to visit https://www.hhs.se/en/about-us/data-protection/ in order to read more and obtain information on your rights related to personal data.

○ I accept the terms stated above

Imagine that you have purchased a piece of clothing online, especially for your friend's upcoming wedding. You had planned in advance and made sure that the delivery date was ahead of the wedding day.

When the package arrives you notice that the clothing is in the wrong size. You go to the website you had ordered it from and read that you can chat with their customer service through the website's chat function.

You read that the way to contact customer service is to chat with their automated chatbot.

You enter the chat function and start to converse with the automated chatbot to firstly, report and return the clothing and secondly, to make sure the new clothing is delivered before the wedding day.

Chatbot: Hello. Thank you for contacting the Automated Customer Service Center. Please explain why you contacted us today.

Your Response: I got the wrong size of the product delivered.

Chatbot: I am very sorry to hear that. Which one of your orders does this concern?

Your Response: Order number 847650.

Chatbot: Okay, we are sorry to hear that you got the wrong product. If you want to return the product I can send you the return documents to fill in. Would you also like me to make a new order for the right size?

Your response: Yes, I would like a new order with the right size.

Chatbot: Super! I'll make the order for the right size. We anticipate that the delivery will take around 4-5 days. Is that all I can help you with today?

Your response: No, I need the clothing delivered tomorrow, as I intend to wear it for a wedding.

Chatbot: I understand, I'll add the express delivery option free of charge! The package will now be delivered to your doorstep by tomorrow morning. Is that fine with you?

Your response: Yes.

Chatbot: Great! Is there anything else I can do for you today?

Your Response: No

Chatbot: Awesome! Happy to help and thank you for contacting customer service. Have a good day!

→

Imagine that you have purchased a piece of clothing online, especially for your friend's upcoming wedding. You had planned in advance and made sure that the delivery date was ahead of the wedding day.

When the package arrives you notice that the clothing is in the wrong size. You go to the website you had ordered it from and read that you can chat with their customer service through the website's chat function.

You read that the way to contact customer service is to chat with a customer service employee.

You enter the chat function and start to converse with the customer service employee to firstly, report and return the clothing and secondly, to make sure the new clothing is delivered before the wedding day.

Employee: Hello. Thank you for contacting the Customer Service Center. Please explain why you contacted us today.

Your Response: I got the wrong size of the product delivered.

Employee: I am very sorry to hear that. Which one of your orders does this concern?

Your Response: Order number 847650.

Employee: Okay, we are sorry to hear that you got the wrong product. If you want to return the product I can send you the return documents to fill in. Would you also like me to make a new order for the right size?

Your response: Yes, I would like a new order with the right size.

Employee: Super! I'll make the order for the right size. We anticipate that the delivery will take around 4-5 days. Is that all I can help you with today?

Your response: No, I need the clothing delivered tomorrow, as I intend to wear it for a wedding.

Employee: I understand, I'll add the express delivery option free of charge! The package will now be delivered to your doorstep by tomorrow morning. Is that fine with you?

Your response: Yes.

Employee: Great! Is there anything else I can do for you today?

Your Response: No

Employee: Awesome! Happy to help and thank you for contacting customer service. Have a good day!

How satisfied are you with the service interaction?

1=Very dissatisfied	2	3	4	5	6	7	8	9	10=Very satisfied
○	○	○	○	○	○	○	○	○	○

How well did the service meet your expectations?

1=Not at all	2	3	4	5	6	7	8	9	10=Totally
○	○	○	○	○	○	○	○	○	○

How likely is it that you remember the interaction with the company as satisfactory?

1=Very unlikely	2	3	4	5	6	7	8	9	10=Very likely
○	○	○	○	○	○	○	○	○	○

Please rate the customer service agent with respect to accessibility if you need help

1=Poor performance	2	3	4	5	6	7	8	9	10=Good performance
○	○	○	○	○	○	○	○	○	○

Please rate the customer service agent with respect to helpfulness

1=Poor performance	2	3	4	5	6	7	8	9	10=Good performance
○	○	○	○	○	○	○	○	○	○

Please rate the customer service agent with respect to friendliness

1=Poor performance	2	3	4	5	6	7	8	9	10=Good performance
○	○	○	○	○	○	○	○	○	○

How realistic do you think the interaction was?

1=Very unrealistic	2	3	4	5	6	7	8	9	10=Very realistic
○	○	○	○	○	○	○	○	○	○

What year were you born?

What gender do you identify as?

Female	Male	Other	Prefer not to say
○	○	○	○

What is the highest level of education you have completed?

Less than high school	High school graduate	Some college	College degree	Doctorate
○	○	○	○	○

Please select the number eight

15	28	39	8	77
○	○	○	○	○

Handelshögskolan i Stockholm

We thank you for your time spent taking this survey.
Your response has been recorded.

7.2 Appendix 2