SCHOOL OF
ECONOMICS AND
MANAGEMENT

‘Perceived risks and benefits of artificial intelligence (AI): A behavioral economics approach’

By

Vilgot Kraft

Master’s Programme in Economics
NEKN01
Master’s Thesis (15 credits ECTS)
June 2023
Supervisor: Roel Van
Veldhuizen

Abstract

This paper aims to analyze individuals' perceptions of artificial intelligence (AI) and how they perceive the potential risks and benefits of AI. This study examines how information treatment effects change people's attitudes toward AI. To analyze this, an experimental survey was conducted on April 27, 2023. A total of 120 respondents were obtained through the research platform Prolific and after data cleaning, a total of 114 responses were utilized in the results.

The results indicate that the following variables, income level and concern for AI, negatively impact whether individuals believe AI is harmful or helpful. Individuals also believe that AI will have a negative impact on the governmental sector. On the contrary, a positive relationship is found in self-assessed risk behavior. Furthermore, a positive relationship is observed in a variable that measures if individuals believe that AI can help humans achieve better outcomes than humans working alone. Additionally, a positive attitude toward AI is found in the following sectors: media and entertainment, financial services, financial advice, and transportation. Nevertheless, the study finds no evidence of a gender difference in the perception of AI. The study finds evidence that individuals update their beliefs when exposed to information treatments. However, the information treatment does not generate significant differences in subsequent survey questions.

Keywords: Artificial Intelligence, Behavioral Economics, Framing, Risk Aversion

Table of Contents

List of Equations

List of Tables

List of Figures

1 Introduction

Technology is faster growing than ever in our current technological revolution. The use of technology is broadening in our society and businesses and society as a whole are trying to find suitable applications.

One advancement has been the use of computational power to create human-like thinking, also known as artificial intelligence (AI) (Lu, 2019). The process to create human-like thinking has been long and began in the nineteenth century. But breakthroughs have been made in recent years and in 2022, an AI service named ChatGPT was launched. ChatGPT is now known as one of the fastest-growing consumer applications and reached 100 million active users only two months after its launch (Maslej et al, 2023). Artificial intelligence has raised concerns around the world, about the potential disruptions that it may cause, and in an open letter, world-leading entrepreneurs called for a pause on the development of more advanced artificial intelligence models than GPT-4 (Future of Life Institute, 2023).

Behavioral economics is an interdisciplinary field that examines individuals' decision-making. It challenges the assumptions of classical economics that individuals are rational and fully informed decision-makers, instead behavioral economics recognizes that human behavior is often influenced by biases and other factors (Kahneman, 2003). As such, behavioral economics offers a more nuanced understanding of decision-making.

This thesis will explore humans' relationship with artificial intelligence from a behavioral economics perspective and more specifically analyze how individuals perceive the risks and potential benefits of artificial intelligence. The thesis will compare how individuals perceive risks and benefits in different contexts and sectors. This thesis aims to answer the following research question:

'Are individuals more likely to perceive potential risks associated with artificial intelligence than potential benefits?'

In this study, the research question will be addressed with an experimental survey. The survey consists of various questions that explore respondents' views and familiarity with AI, as well as their concerns and general risk behavior. To enhance the analysis, the survey will utilize an information treatment design, incorporating two different treatments: a negative treatment and a positive treatment of AI. The thesis will investigate how the treatment influences the respondents' views and behavior related to AI. Additionally, the survey explores if the respondents are risk-averse.

The thesis is organized as follows: Chapter 2 introduces behavioral economics and its theoretical background and concepts. Later in this chapter artificial intelligence and its adaptations and concerns are introduced. Chapter 3 explains the methodology and the experimental design used. Chapter 4 presents the results, including demographic data, and statistical analysis. Followed by an in-depth analysis and discussion in Chapter 5. Chapter 6 offers concluding remarks, and a reference list is provided in Chapter 7. Finally, the complete survey is provided in Appendix A.

2 Background

2.1 Behavioral Economics

Behavioral economics is an interdisciplinary field of study that combines principles from psychology, sociology, and economics, it relates human behavior to the allocation of resources. The field does not intend to replace the standard framework of analysis but rather extend it with more realistic psychological foundations to increase the explanatory power of economics (Wilkinson & Klaes, 2018).

2.1.1 Rationality

Rationality is a normative model of decision-making in neoclassical economics, the model aims to describe individuals' behavior and direct how they should behave given certain preferences and restrictions to maximize their utility. However, the concept of rationality is used in many different ways, both in economics and in other fields. For example, acting rationally could just be an individual using reason. Economists usually regard this interpretation as imprecise which has led to a framework of rational decision-making. The framework is built upon two axioms, completeness, and transitivity. Completeness means that individuals have preference ordering across all possible courses of action while transitivity refers to consistency, for example, if an individual prefers A to B and B is preferred to C then A is preferred to C. These two axioms form what is known to economists as rational individuals. However, this model is often extended with the axioms of monotonicity and convexity. Montonistic preferences mean that an individual prefers more of a good rather than less of it, whilst convex preferences are regarded as averages are better than extremes (Klaes & Wilkinson, 2018)

This model of rationality is limited to decisions under certainty, where all individuals have access to the same and complete information. Thus, this model cannot be used to analyze decisions under uncertainty, historically this led to the development of more complex models within behavioral economics (Klaes & Wilkinson, 2018)

In the following subchapters extensions of the rational choice model will be discussed.

2.1.2 Bounded Rationality

Simon (1957) proposes to replace the idea of utility maximization and rationality with a more realistic view of economic behavior which involves humans' cognitive capabilities. Simon (1957) argues that individuals can only make rational decisions within the limits of their cognitive abilities, available information, and the time constraints they face. Thus, he proposes the alternative model, bounded rationality, and satisficing. Bounded rationality is an adaptation of aspirations of success and failure. Individuals' decision-making is influenced by both internal factors such as emotions and values, and external factors such as social norms and cultural influences. Individuals are subjective and use satisficing to make reasonable decisions coherent with their interpretation of the world (Simon, 1957).

Kahneman (2003) presents extensions of bounded rationality and the concept maps of bounded rationality. Maps of bounded rationality are a framework for understanding decision-making that considers the limitations of human cognition. Humans have constrained calculation ability and memory. The framework explores systematic biases and differences between the optimal choices in traditional models and behavior observed in a context (Kahneman, 2003).

According to the distinction proposed by Stanovich and West (2000), human cognition can be categorized into two systems: System 1 and System 2. System 1 is characterized by intuitive, effortless, and unconscious processing, while System 2 involves effortful, slower, and conscious reasoning. Stanovich and West (2000) argue that individual differences in reasoning can be explained by differences in the effectiveness of these two cognitive systems. Some individuals may have a stronger reliance on one system over the other, which can lead to differences in rationality and decision-making. Kahneman (2003) states that there are examples where intuition is associated with poor performance. However, humans can practice intuitive thinking and acquire a higher accuracy by being more familiar with the task. For example, experienced nurses can detect subtle signs of impending heart failure (Kahneman, 2003). Differences in effort distinguish whether a mental process should be assigned to System 1 or System 2, and it is also affected by the individual's perceptions and preferences. Thus, the individual's behavior is connected to their emotions and context (Kahneman, 2003).

Jolls, Sunstein, and Thaler (2000) extend the analysis of behavior and decision-making beyond bounded rationality and introduce limited willpower and bounded self-interest. Limited willpower refers to the tendency for individuals to act against their long-term interests because they lack the willpower to make choices that are objectively better in the long run. For instance, most smokers say that they would not prefer to smoke, and many people pay money to either join a program to help them quit or take supplementary drugs that are said to help them quit. Jolls, Sunstein, and Thaler (2000) argue that people generally recognize their limited willpower but have difficulties in changing their behavior and instead seek ways to mitigate it. For example, joining a pension plan. A pension plan is a special savings arrangement where funds only can be withdrawn after the person reached retirement age, which helps prevent undersaving Jolls, Sunstein, and Thaler (2000). Limited willpower is relevant in situations involving long-term consequences, such as criminal behavior. However, it may not be applied universally and therefore it is important to consider the specific context where decision-making is discussed (Jolls, Sunstein, and Thaler, 2000).

Bounded self-interest refers to the concept that people are bounded by social factors such as empathy, altruism, and social norms. Jolls, Sunstein, and Thaler (2000) argue that individuals are not perfectly rational and always act in their self-interest when it comes to decision-making. Instead, they are influenced by these factors which conflict with their self-interest. For example, people may donate money to charity, even though they are sacrificing their own financial gain because they value helping others or because they might feel social pressure to do so.

Posner (1998) argues that the concept of bounded rationality is too vague and imprecise to analyze decision-making. He suggests that the concept is often used to justify assumptions about human behavior that are not supported by empirical evidence. Posner (1998) states that bounded rationality is unnecessary as it is already incorporated in traditional economic models through the utility function, which function allows for irrational or inconsistent behavior. Posner's criticism is not universally accepted, and the research Daniel Kahneman conducted together with his colleague Amos Tversky was acknowledged with the award of the Nobel Memorial Prize in Economic Sciences in 2002 (Smith, 2002).

2.1.3 Prospect Theory

Kahneman and Tversky (1979) introduce Prospect Theory as a method to explain decisions made under risk which violates the axioms of rationality and expected utility theory.

Kahneman and Tversky (1979) argue that individuals exhibit a certainty effect where they overvalue certain outcomes in comparison to probable outcomes. Kahneman and Tversky (1979) point out that when individuals face outcomes that are either certain or uncertain, they are risk-averse when the outcome involves gains. Opposite, when the outcome involves losses, individuals are more likely to choose the uncertain outcome and thus become risk-seeking. Thus, the certainty effect leads to different risk preferences depending on whether the outcome involves gains or losses (Kahneman and Tversky, 1979). Secondly, decision-making is also influenced by the isolation effect. Kahneman and Tversky (1979) argue that individuals tend to disregard commonalities among options and rather focus on the unique features. Essentially, the isolation effect can be utilized to frame an individual's decision-making. Depending on how the options are presented, they can be framed either positively or negatively, and this framing can cause people to perceive and evaluate them differently (Kahneman and Tversky, 1979). Thirdly, the reflection effect is a concept within decision-making. It means that individuals' choices are mirror images of each other when faced with outcomes that are positive or negative. This idea suggests that individuals' decision-making process is influenced by how the options are presented and how they emotionally perceive the outcomes, rather than just the objective characteristics of the options themselves (Kahneman and Tversky, 1979).

2.1.4 Loss Aversion and Status Quo Bias

Prospect Theory suggests that a decision-maker's wealth reference point is a basis for determining the value of an option, rather than the final state of wealth. Kahneman and Tversky (1979) derive a hypothetical value function that is characterized by the following: it is defined on deviations from the reference point, generally concave for gains and convex for losses, and steeper for losses than for gains. An illustration of the value function is shown below in Figure 2.1.

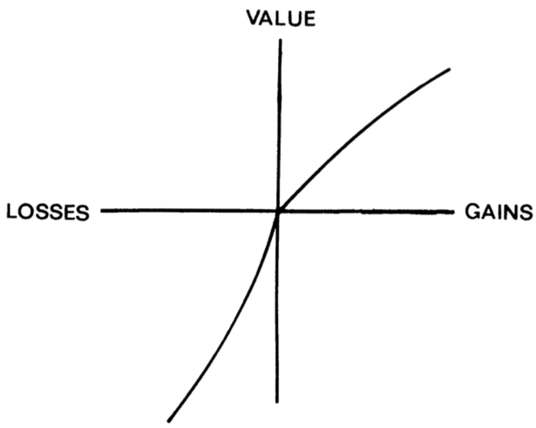

Figure 2.1 A hypothetical Prospect Theory value function (Kahneman and Tversky, 1979).

This is known as loss aversion, individuals experience a stronger negative emotional response to losses than to gains, hence the steeper gradient. Kahneman and Tversky (1979) argue that loss aversion can help explain why people resist changes and prefer to maintain the status quo, even though changes might be beneficial. Because individuals tend to focus on the potential loss rather than the potential gain. Thus, they are more likely to be risk-averse when faced with a probability of a loss. Additionally, individuals anchor to their current state (status quo), and any deviation from that status quo is viewed as a loss (Kahneman and Tversky, 1979).

Samuelson and Zeckhauser (1988) find evidence for the cognitive bias known as status quo bias. In their experiment, respondents were presented with a hypothetical inheritance and were asked to choose between different investment options, some with a neutral description, and some options were designed as status quo. The options designated as status quo were significantly more likely to be chosen than the alternatives, and as the number of presented alternatives increased so did the advantage of the status quo. The findings of Samuelson and Zeckhauser (1988) suggest that individuals are more likely to maintain the status quo even though there are better options available, which they refer to as the status quo bias.

Hartman et al. (1991) also find evidence for the status quo bias; their study investigates the existence of the status quo in consumer decision-making. Respondents were presented with options that were familiar, status quo, and unfamiliar but potentially better. The study showed that participants were significantly more likely to choose the status quo alternative even though the unfamiliar option was objectively a better option. The study suggests that status quo bias is a pervasive cognitive bias that affects consumer decision-making, and the bias might lead to suboptimal outcomes as consumers deflect objectively better options.

2.1.4.1 Endowment Effect

Another anomaly is found by Thaler (1980), in his experimental study, respondents are asked to value buying and selling prices of various tasks and goods. Thaler (1980) finds a noticeable difference in buying and selling prices, which could be explained by income effects and transaction costs. However, Thaler (1980) emphasizes that all costs are opportunity costs and should be treated as equivalent to out-of-pocket costs. As argued by Kahneman and Tversky (1979) losses and gains tend to be valued differently, and individuals have a value function with the characteristics of Figure 2.1. Given these characteristics, Thaler (1980) states that individual out-of-pocket costs will be viewed as losses, and opportunity costs will be viewed as foregone gains. Thus, out-of-pocket costs will be given more weight. Thaler (1980) describes this underweighting of opportunity costs as the *endowment effect.* Thaler (1980) states that the endowment effect can explain why individuals are more reluctant to give up goods they already own even if it would be objectively better to do so.

Knetsch (1989) finds evidence for the endowment effect, that losses from a reference position are valued differently than corresponding gains. His findings are consistent with the suggestions by Thaler (1980), however, Knetsch (1989) argues that the endowment effect may have important welfare implications, as it can create market inefficiencies and distortions because of inefficient resource allocations.

2.1.5 Framing and reference point

Kahneman and Tversky (1986) distinguish between two phases in the decision-making process, framing, and evaluation. In the framing phase, the individual is analyzing the decision problem and tries to identify effective responses, contingencies, and outcomes. The individual is influenced by how the choice is presented. The individual is also influenced by norms, biases, and expectations. It is in the evaluation phase where the individual assesses the prospects and selects the one with the highest value. Framing theory suggests that an individual's behavior can either be compared by detecting dominant choices or by comparing their values. Kahneman and Tversky (1986) find evidence for the framing effect in simple gambles with numerical probabilities and monetary outcomes.

Kahneman and Tversky (1986) also find evidence for the importance of reference points. The reference point is one aspect of framing and is the initial point from which gains and losses are evaluated. Thus, it serves as a baseline for comparison and can be influenced framing. Kahneman and Tversky (1986) show in one of their problems that the reference point can be shifted by simple labeling of a problem.

Kahneman and Tversky (1986) remark that even though the analysis of framing effects can provide greater insight into decision-making, their relaxed analysis of framing effects is complicated. Factors such as language, context, and the nature of display influence the analysis making it difficult to provide a complete and formal theory of framing. For example, in their research, they do not account for failures of transitivity and invariance of choices (Kahneman and Tversky, 1986).

2.1.6 Availability Heuristic

Another judgmental heuristic introduced by Kahneman and Tversky (1974) is availability. Availability is a cognitive bias in which individuals overestimate the frequency or likelihood of incidents based on how easily they can retrieve examples or instances of those incidents. Availability occurs when individuals rely more on information that is easily accessible and vivid in their memory, rather than considering other aspects and relevant information (Kahneman and Tversky 1974).

Kahneman and Tversky (1974) note that both salience and familiarity can impact the bias, for example, seeing an incident may have a greater impact than reading about an incident in the paper. Familiarity can impact the bias as people may assess the risk of having a heart attack by evaluating the occurrences among one's acquaintances.

The occurrence of availability bias may induce errors in judgment and decision-making, as individuals may rely on incomplete information rather than consider a variety of sources and critically evaluate information before making decisions.

2.1.7 Risk Aversion

Risk aversion is a behavioral tendency in expected utility theory that makes people choose a certain outcome with a lower expected payoff rather than choosing a risky outcome with a higher expected payoff. Risk-averse individuals tend to prefer certainty over uncertainty in decision-making and are therefore more likely to avoid risks. Risk aversion is often divided into two types, absolute risk aversion, and relative risk aversion. Absolute risk aversion refers to the propensity to avoid risks regardless of their expected payoff, while relative risk aversion refers to the propensity to avoid risks depending on the individual's current situation and wealth (Stefánsson and Bradley, 2019).

2.2 Artificial Intelligence (AI)

Kurzweil (1985) states that the development of intelligent machines began in the early nineteenth century, with ideas of a machine that could think, such as the 'Analytical Engine' designed by Charles Babbage who contemplated in the field. During that time, those concepts lacked the ability to deeply influence or have a significant impact. Later in 1956 at a Dartmouth conference, John McCarthy came to name intelligent machines as artificial intelligence and the field gained public attention and during the years there has been great controversy regarding what constitutes AI (Kurzweil, 1985).

Yang Lu has defined AI as "Any theory, method, and technique that helps machines (especially computers) to analyze, simulate, exploit, and explore human thinking process and behavior can be considered as AI." (Lu, 2019, p.1). Another definition made by IEEE (Institute of Electrical and Electronics Engineers) is "Artificial Intelligence is that activity devoted to making machines intelligent, and intelligence is quality that enables an entity to function appropriately and with foresight in its environment" (Institute of Electrical and Electronics Engineers, 2019, p.1).

AI is a field multidisciplinary and interdisciplinary field of study that is characterized by computational systems designed to execute tasks that ordinarily required human intelligence. These tasks are, for example, understanding natural language, decision-making, and problem-solving. Although the definition of AI varies, the core is a notion of intelligent systems. These systems can simulate human behavior and abilities and thus enable computers to execute more complex and automated tasks (Došilović, Brčić, and Hlupić, 2018; Mata et al., 2018; Tan and Lim, 2018).

In the recent decade, there has been a significant increase in the interest in AI and its applications, for example the number of AI publications in the world has increased from 200,000 in 2010 to approximately 500,000 in 2021 whilst the number of AI Journal publications increased from approximately 100,000 in 2010 to 293,000 in 2021 (Maslej et al., 2023). Similar trends can also be seen in investment activities in AI, for example, the private investment activities in 2022 has 18-fold since 2013, even though there is a significant decrease in private investment activity between 2021 and 2022. In 2021, private investments in AI reached its peak at a total of 125.36 billion dollars but decreased to 91.86 billion dollars in 2022 (Maslej et al., 2023).

The interest in AI is widespread and there are numerous adaptations in the society. The following section provides a few examples of AI implementation, though it is not intended to be a comprehensive list. As India is sensitive to rainfall, AI algorithms have been implemented to analyze and predict rainfall and seasonal variations (Dash, Mishra, and Panigrahi, 2018). The automobile industry is working towards self-driving cars and in recent years autopilot systems (AI) have been implemented and tested (Stilgoe, 2017). In the medical field, AI is developing quickly, and implementations of an AI system have been done in China to support treatment recommendations for patients who have lung cancer (Liu et al., 2018). In the finance industry, particularly in Financial Technology (FinTech) companies, AI has been implemented to perform various tasks such as analyzing big data and credit risks, providing customer assistance, and mitigating money laundering (Sharbek, 2022).

2.2.1 Chatbots and other popular AI systems

In 2022, AI systems such as DALL-E and ChatGPT gained popularity. These systems can perform numerous tasks such as text manipulation, analysis, image generation, and question answering. Despite their capabilities, these systems may generate hallucinatory responses with confidence, which could be either incoherent or untrue. Making it difficult to depend on them for critical applications. Despite this, ChatGPT reached 100 million monthly active users two months after release, thereby establishing a new record as the fastest-growing consumer application in history, for comparison it took TikTok about nine months, and Instagram about two and a half years to gain the same amount of users (Hu, 2023; Maslej et al., 2023).

2.2.2 Concerns

The ongoing development within the field of AI is of great interest to policymakers, industry leaders, researchers, and the public. According to Maslej et al (2023), the continuous improvement of AI will lead to a growing integration into society. The potential for massive disruption sets high demands on society's desired evolution and prosperity of AI.

In an open letter sent on March 22, 2023, co-signed by Tesla CEO Elon Musk and Apple co-founder Steve Wozniak, calling for a pause on giant AI experiments that are more advanced than GPT-4 (Future of Life Institute, 2023). Musk and Wozniak are expressing concerns related to privacy issues, potential disruption of industries, lack of governance and audit, and potential black-box models. Black-box models refer to systems that have emergent capabilities which are not known by the developers. The open letter is currently signed by 27,567 people, including numerous industry leaders and researchers. A comprehensive list can be found on the web (Future of Life Institute, 2023).

3 Methodology

3.1 Data Collection

All data in this study is obtained from an experiment done online with the online tool, Google Forms. The data from the experiment is collected solely for this thesis. Before the experiment, on 26 April 2023, a pilot study was done with seven participants to confirm that each question was intuitive and understood. Thereafter, the experiment was conducted on 27 April 2023, with a survey panel obtained through Prolific. Prolific is an online tool that helps researchers obtain high-quality data in online experiments and surveys by connecting them with respondents across the world (Prolific, 2022). Respondents from the Nordic countries (e.g., Sweden, Norway, Denmark, Iceland, and Finland) were allowed to participate, the number of available respondents that had been active on Prolific in the recent 90 days was 674, among these 674 participants, the study collected 120 responses. The study was available the 27 April 2023. The participants were incentivized to participate with a payment of 1.3 GBP. There were no boundaries to conduct the survey and all data remained anonymous, hence no personal information was disclosed by participants. After successful submission, answers were manually reviewed before payment was approved. The complete survey is provided in Appendix A.

3.2 Experiment Design

The Google Forms survey consisted of 21 different questions, of which each respondent was required to answer 20 of these. The initial question of the survey asked participants to provide their unique Prolific ID. The following eight questions asked respondents to grade their current beliefs and familiarity with artificial intelligence on a five-point Likert scale, as well as their usage of AI services and chatbots. Additionally, participants were also asked to self-assess their willingness to take on risks in general, a question copied from the German Socio-Economic Panel (SOEP) survey. Such a question has been commonly employed in previous research and shown to be predictive of actual risk behavior (Ding, Hartog, and Sun, 2010; Dohmen et al., 2005).

In section 2 of the survey, the respondent is asked to read the section preamble carefully, which discloses the answer to question 10. This question was designed as an attention check to assess respondents' attentiveness and to prevent hasty responses (Krosnick, 1991). As recommended by Haaland, Roth, and Wohlfart (2023) it is important to disclose why attention checks are being used to mitigate concerns and negative reactions, they also argue that it is important that attention checks are intuitive, not cognitively demanding, and straightforward.

Question 11 was included as a randomizing mechanism, Specifically, respondents were asked if they were born in an odd or even month and were directed to either question 12 or 13 accordingly. This mechanism was not disclosed to the respondents and thus, each group was only able to answer one of the questions. The purpose of questions 12 and 13 is to test whether individuals respond to the treatment effects and change their attitude towards AI. The questions are priming the respondent in a positive respectively negative way and thereafter repeat question four. Such information treatment may have side-effects namely uncertainty and emotional responses. However, active control has benefits when it comes to studying the causal effect of expectations on behavior. The information intervention in the questions was general and quantitative and aims to change the perception of the social norm. Therefore, the information intervention aims to measure posterior beliefs and measure the robustness of the prior beliefs (Haaland, Roth, and Wohlfart, 2023).

Questions 14 and 15 asked the respondents to evaluate two hypothetical questions, question 14 asked the respondent to choose between an AI advisor and a human advisor when conducting an investment, and question 15 asked if the participant would like to participate in a lottery with equal probabilities of winning and losing 10 euro. Hypothetical questions have a potential disadvantage as they might not be able to predict actual behavior, to mitigate this one can combine the survey experiment with a field experiment or use real money in the experiment. However, hypothetical questions have the advantage that they possibly can measure attitudes at a relatively low cost (Dohmen et al., 2005). Kahneman and Tversky (1979) argue that while hypothetical questions have potential drawbacks, they are still valuable for exploring complex decision-making scenarios that cannot be observed in field or laboratory experiments. Field experiments also have limitations, such as difficulties in accurately measuring utilities and probabilities. Whilst laboratory experiments can obtain precise measures of utility and probability from choices, they lack the ease of interpretation and are restricted in the generalizability of results. Kahneman and Tversky (1979) suggest that hypothetical questions may be useful if participants know how they would behave in real situations and if they have no special reason to hide or misrepresent their true preferences. Therefore, in the preamble for questions, 14 and 15 participants are reminded and asked to imagine that they are facing the actual problem and indicate how they would act in real life. The preamble also states that there is no correct answer to the questions and that the participants' answers will remain anonymous.

To reduce the possibility of stereotype threats in the study, the study concluded with six demographic questions regarding age, sex, income, education level, occupation, and sector affiliation. Additional steps such as using inclusive language and alternatives were also used throughout the experiment (Steele and Aronson, 1995).

3.3 Pilot Study

In the pilot study, 57.1% of the sample failed the attention check in question 10. Therefore, the phrasing was changed to underline the importance of reading the preamble carefully and to emphasize that an answer was provided in the preamble. The updated preamble is presented below:

"Please read this section carefully as you will be asked two questions and one answer will be provided in the text. The first question is about a problem. In questionnaires, sometimes there are participants who do not carefully read the questions, this could compromise the study. To show that you read the questions carefully, please enter number five (5) as the answer to the next question, what is your favorite number? The second question will ask you to state if you are born in an 'odd' or 'even' month."

4 Results

In this section, the results of the experiment are presented. First, general information about the experiment is provided. Then, demographic data is presented. Finally, detailed results of the remainder of the experiment are described and presented.

Out of the 120 respondents, six failed the attention check in question 10 and had their responses invalidated and excluded. Therefore, the data in this study is based on 114 respondents. The data was recoded to simplify the analysis process. For example, to simplify the analysis of question 4, the original response options were recoded on a scale that ranged from -2 (Mostly harmful) to 2 (Mostly helpful). This was done similarly for the other questions that needed to be recoded into numerical values for interpretation.

4.1 Demographic Data

The demographic data is provided in Table 4.1. More than a majority of the respondents were between 21 – 40 years old (78.9%), 67.5% of the sample were male respondents, and 61.3% of the sample had a completed bachelor's degree or higher. At least 11 sectors were represented in the sample, and 29.9% of the respondents did not identify with any of the sectors listed.

Table 4.1 Demographic data

Sample characteristics, N=114		***n***	***%***
Age Range			
	11-20	1	0.9
	21-30	47	41.2
	31-40	43	37.7
	41-50	13	11.4
	51-60	7	6.1
	71-70	3	2.6
Sex			
	Female	36	31.6
	Male	77	67.5
	Non-binary	1	0.9
Occupation			
	Seeking employment	14	12.3
	Employed	52	45.6
	Student	29	25.4
	Entrepreneur	7	6.1
	Unemployed	7	6.1
	Retired	3	2.6
	Other	2	1.8
Education			
	Lower secondary education	4	3.5
	Upper secondary education	30	26.3
	Vocational education	8	7.0
	Bachelor's degree	42	36.8
	Master's degree	21	18.4
	Doctoral degree	7	6.1
	Other	1	0.9
	Prefer not to say	1	0.9
Sector			
	Financial operations, business services	9	7.9
	Human health and social work	8	7.0
	Education	16	14.0
	Manufacturing	6	5.3
	Trade	2	1.8
	Public administration	2	1.8
	Information and communication	26	22.8
	Personal and cultural services	3	2.6
	Transport	3	2.6
	Agriculture, forestry, and fishing	4	3.5
	Accommodation and food services	2	1.8
	Other	33	28.9
Income			
	< 10,000	25	21.9
	10,001 - 30,000	32	28.1
	30,001 - 50,000	22	19.3
	50,001 - 70,000	10	8.8
	> 70,001	10	8.8
	Prefer not to say	15	13.2

Note: Education refers to the highest level of education completed. Manufacturing is a collective term encompassing manufacturing, mining, and quarrying, energy, and environment. All income figures are in euros. The total number of respondents is 114.

4.2 Experiment Results

In Figure 4.1, respondents' self-assessed familiarity with AI is shown. Roughly 75% have selected moderately familiar or higher, indicating that they have some level of familiarity with AI. This is also shown in Figure 4.2, where almost 50% of the respondents have used ChatGPT. ChatGPT quickly gained in popularity at the end of 2022 (Maslej et al., 2023). As seen in the figure, none of the other AI services have such a widespread userbase. Some respondents added options such as MyAI on Snapchat, and ChatPDF.

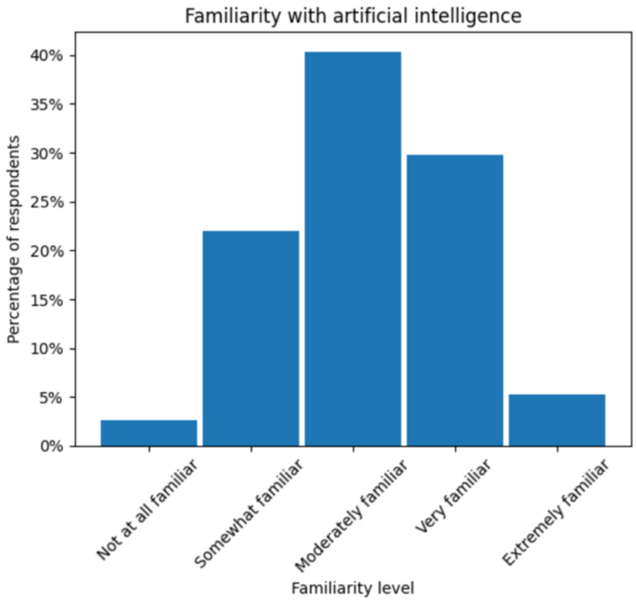

Figure 4.1 Familiarity with artificial intelligence

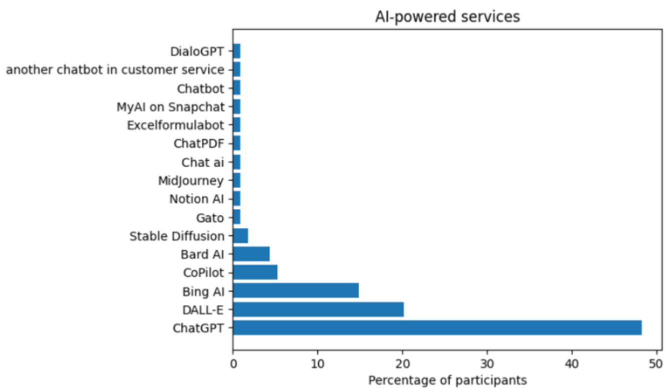

Figure 4.2 Usage of AI-powered services

Table 4.2 displays the data for questions 5 and 6. In question 5, more than 50% of the participants perceive that AI to some extent will have a positive impact in the following sectors, healthcare, financial services, financial advice, Retail, Transportation, and Manufacturing. Meanwhile, the respondents perceived that the sectors with the most negative impact will be media and entertainment, and government. However, government is also the sector where respondents were most uncertain about the effects, which can be compared to healthcare, the sector with the lowest uncertainty.

In question 6, more than 50% of the respondents had to some extent a positive view of AI's impact on decision-making in the following contexts, financial decisions, healthcare decisions, and personal life decisions. The view on financial decisions, and healthcare decisions is consistent with the perceptions in question 5. Notably, 32.5% of the respondents had to some extent a negative view of AI's impact on employment decisions, while 34.2% were uncertain about AI's impact, and 33.4% had to some extent a positive view on employment decisions, indicating a polarized view in this context.

To calculate the average in Table 4.2, the responses have been recoded from -2 (Mostly negative) with one-unit increments to 2 (Mostly positive).

Table 4.2 Perceptions of artificial intelligence

Perceptions	***Mostly negative***	***Somewhat negative***	***Neither positive nor negative***	***Somewhat positive***	***Mostly positive***	***Average***
Q.5: In the sectors specified below, do you think artificial intelligence will have a positive or negative impact?	*Percentage of respondents (%)*					
Healthcare	1.8	11.4	11.4	40.4	35.1	0.96
Education	8.8	22.8	18.4	30.7	19.3	0.29
Financial services	4.4	8.8	21.1	47.4	18.4	0.67
Financial advice	4.4	10.5	25.4	44.7	14.9	0.55
Retail	4.4	14.0	21.9	38.6	21.1	0.58
Transportation	1.8	6.1	18.4	50.0	23.7	0.88
Manufacturing	4.4	1.8	21.9	40.4	31.6	0.93
Legal	12.3	24.6	26.3	24.6	12.3	0.00
Media and entertainment	11.4	28.9	21.1	24.6	14.0	0.01
Government	19.3	23.7	33.3	14.9	8.8	-0.30
	Strongly disagree	***Disagree***	***Neither agree nor disagree***	***Agree***	***Strongly agree***	***Average***
Q.6: To what extent do you agree or disagree with 'artificial intelligence will help people make better decisions in the following contexts?'	*Percentage of respondents (%)*					
Financial decisions	1.8	14.9	30.7	42.1	10.5	0.45
Healthcare decisions	3.5	17.5	22.8	41.2	14.9	0.46
Legal decisions	8.8	20.2	27.2	33.3	10.5	0.17
Employment decisions	12.3	20.2	34.2	28.1	5.3	-0.06
Personal life decisions	5.3	19.3	23.7	43.9	7.9	0.30

As Google Forms does not have the option to randomize questions among respondents, and even if they did, the randomizing mechanism would not be transparent. Therefore, a manual randomizing mechanism was applied, based on whether respondents were born in an odd or even month. Table 4.3 shows that the participants were slightly skewed towards even months, with 63 participants receiving the positively framed question and 51 participants receiving the negatively framed question.

Table 4.3 Randomizing mechanism

Randomizing mechanism		***n***	***%***
Q.11: Are you born in an 'odd' or 'even' month?			
	Odd	63	55.3
	Even	51	44.7
	Total	114	100

Table 4.4 summarizes the results of question 4. The responses to question 4 are also shown in Figure 4.3, which indicates that roughly 50% of the respondents believe that AI is somewhat helpful. This is confirmed by the data in Table 4.4, as both the mean and the 95% confidence interval suggest a moderate level of perceived helpfulness for AI.

Table 4.4 Helpful or harmful

Helpful or harmful	***n***	***Mean***	***S.D.***	***95% C.I.***
Q.4: Do you think artificial intelligence is mostly helpful or mostly harmful to people?				
	114	0.614	0.945	0.439 – 0.789

Note: S.D. represents the standard deviation, and 95% C.I. represents the 95% confidence interval.

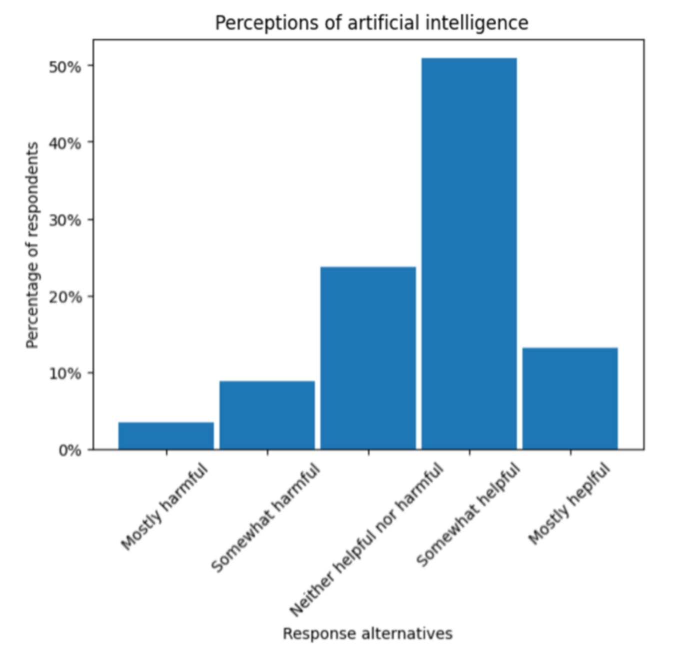

Figure 4.3 Perceptions of artificial intelligence

A difference of means test was conducted to test whether the groups responded differently to the framing of questions 12 and 13. The following null hypothesis was used:

$$H_0: difference\ of\ means = 0$$

Equation 4.1

As shown in Table 4.5, the t-test for the null hypothesis specified in equation 4.1 is significant, both the two-tailed and one-tailed p-values are significant at the 0.01 level, indicating that there is an observed difference between the two framings. Notably, the two 95% confidence intervals are not overlapping, emphasizing the difference between the different framing of the questions.

Table 4.5 Difference of means

Differences of means	***n***	***Mean***	***S.D.***	***95% C.I.***
Q.12-13: Positive and negative framing, "Do you think artificial intelligence is mostly helpful or mostly harmful?"				
Positive	63	0.730	0.807	0.527 – 0.933
Negative	51	0.196	1.11	(-0.117) – 0.509
Degrees of Freedom		112		
T-statistic		2.965		
Two-tailed p-value		0.00370(**)		
One-tailed p-value		0.00185(**)		

Note: Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001. S.D. represents the standard deviation, and 95% C.I. represents the 95% confidence interval.

In question 7 respondents were also asked about their concerns about artificial intelligence, which is shown in Figure 4.4, a small group of respondents was not concerned at all, roughly 7%. However, most respondents expressed some concern, and approximately 25% of the respondents were quite concerned or very concerned.

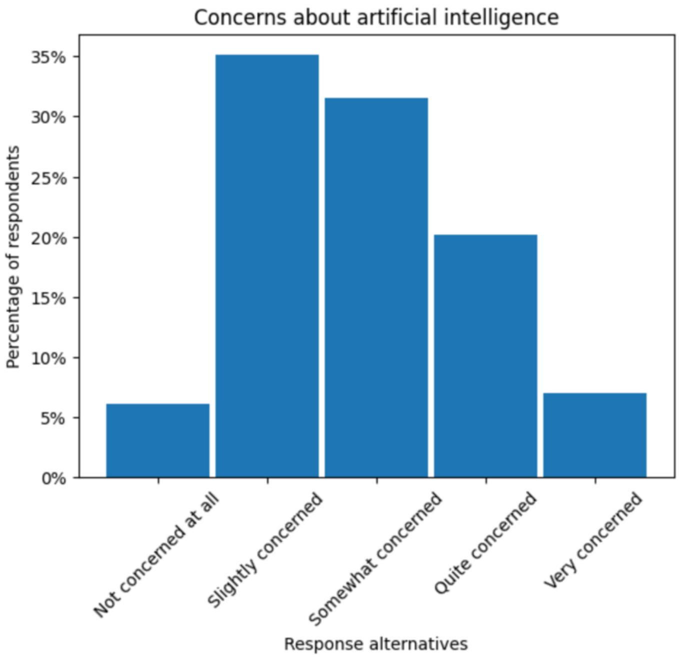

Figure 4.4 Concerns about artificial intelligence

In question 8, the respondents were asked "To what extent do you agree or disagree with 'it is possible for artificial intelligence to help humans achieve better outcomes than human beings working alone'?" these responses are shown in Figure 4.5. This clearly shows that there is a positive attitude towards AI and that AI has potential benefits for humans, less than 10% disagree or strongly disagree with this statement.

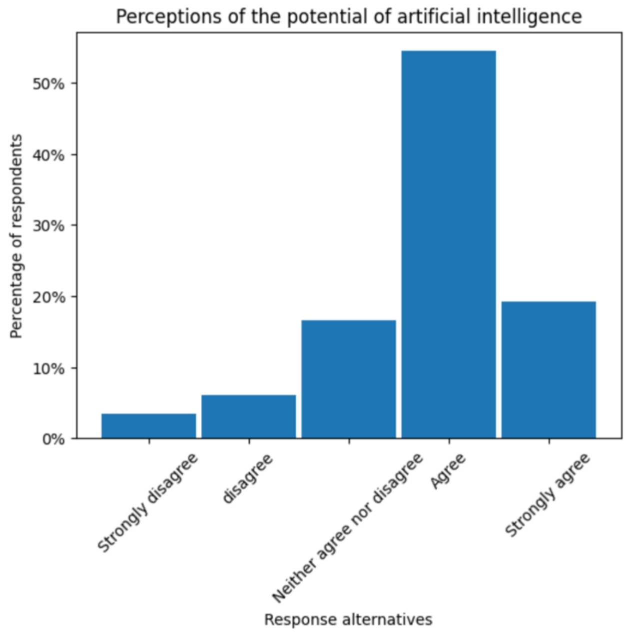

Figure 4.5 Perceptions about the potential of artificial intelligence

Figure 4.6 illustrates the respondents' self-assessed risk appetite in question 9 where 0 indicates 'unwilling to take risks' and 10 indicates 'fully prepared to take risks'. Option 5 is the middle alternative, and as captured by the data in Table 4.6, the sample is slightly skewed to a higher willingness to take risks. But the 95% confidence interval is still capturing the middle alternative.

Table 4.6 Risk appetite

Risk appetite		*Mean*	*Median*	*S.D.*	*95% C.I.*
Q.9: How do you see yourself: Are you in general a person who takes risk or do you try to evade risks?					
	Sample	5.368	6.0	2.267	4.952 – 5.784

Note: S.D. represents the standard deviation, and 95% C.I. represents the 95% confidence interval.

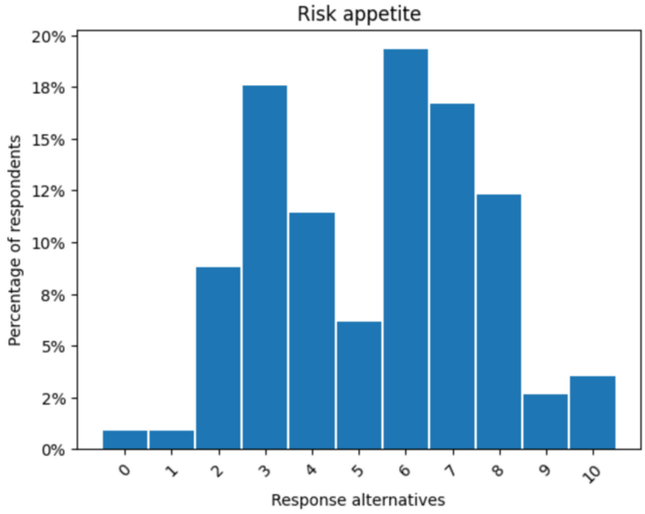

Figure 4.6 Risk appetite

Figure 4.7 shows the results of question 14. This question asked respondents to choose whether they wanted a human advisor or an AI advisor when conducting an investment. In this hypothetical question, the AI advisor had a fixed probability of success of the investment of 50% and the human advisor had a success probability ranging between 60 - 40%. As the human advisor's success probability decreased, the preference for the AI advisor increased. When they had an equal success probability of 50%, 62 (54.4%) of the participants preferred the human advisor while 52 (45.6%) preferred the AI advisor. Two respondents showed inconsistency in their preferences. One of them when the success probability of the investment changed from 60% to 57.5% and the other of them when the success probability changed from 42.5% to 40%. At the 60% level, one of the two respondents preferred the AI advisor. But at the 57.5% level, the respondent changed their preference towards the human advisor. Similarly, at the 42.5% level, one of the two respondents preferred the AI advisor but at the 40% level, the respondent changed their preference towards the human advisor. These inconsistencies are illustrated more in detail in Figures 4.8 and 4.9.

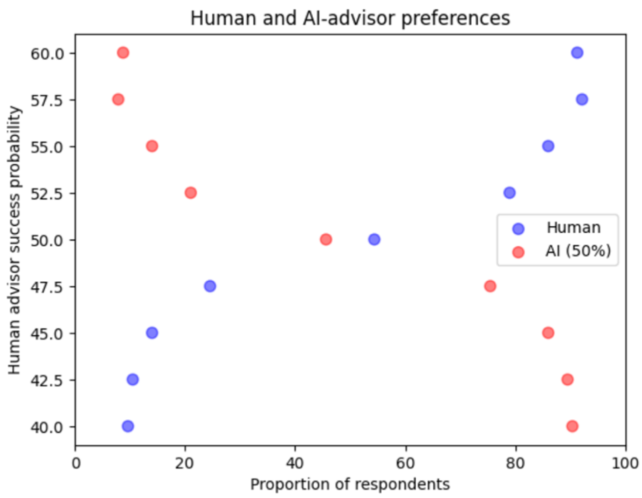

Figure 4.7 Advisor preferences

In the last question 15, respondents were asked if they wanted to participate in a hypothetical lottery. The question was stated in the following way "Would you like to participate in a lottery where you have a 50% chance to win 10 EUR and a 50% chance to lose 10 EUR?". 51 (44.7%) participants stated that they would like to participate, and 63 (55.3%) participants stated that they would not like to join. The expected utility of not joining was zero (0), and the expected utility for joining was also zero (0). This indicates that more than half of the participants are risk averse.

Figure 4.8 illustrates the results of the positive treatment effect for question 14. In comparison with Figure 4.7, the positive framing does not increase the tendency for the AI advisor; rather the opposite occurs. There is a slight increase in preference for the human advisor at the critical point (50%) success probability.

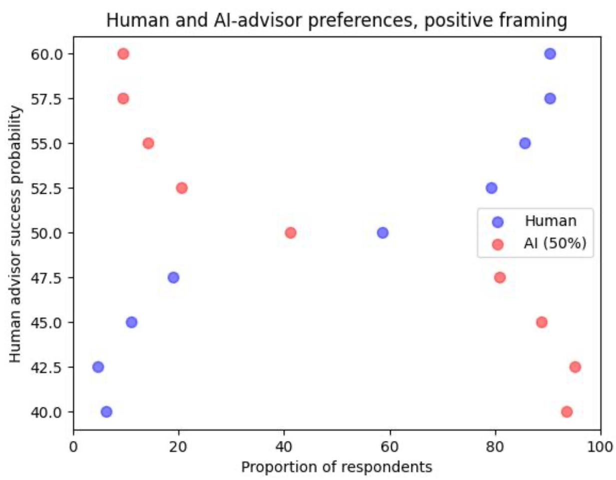

Figure 4.8 Advisor preferences, positive framing

Figure 4.9 illustrates that the respondents who received a negative treatment, have a higher preference for the AI advisor at the critical point than for the human advisor. However, the preference is not significant, but there is an increase from the total dataset.

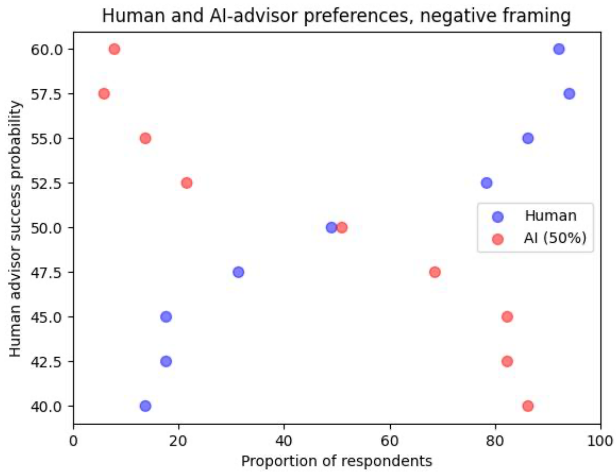

Figure 4.9 Advisor preferences, negative framing

Table 4.7 presents the responses to question 15, the lottery. The table is divided between the two treatment groups. The positive treatment group is risk-averse to a larger extent than the negative treatment group.

Table 4.7 Framing effect, lottery

Framing effect, lottery	***n***	***Yes***	***No***
Q.15: Would you like to participate in a lottery where you have 50% chance to win 10 EUR and 50% chance to lose 10 EUR?		*n, (percentage of respondents)*	
Positive framing	63	26, (41)	37, (59)
Negative framing	51	25, (49)	26, (51)

4.3 Regression Analysis

The data from the experiment was recoded in order to analyze it with ordinary least squared (OLS) regression. In Table 4.8, the question's regression names are presented, which question each variable corresponds to, and how it was recoded. Note that, there is only one (1) dependent variable, the variable 'Familiar' was tested as an independent variable but did estimate a strong r-squared.

Table 4.8 Regression variables

Regression variables		***Question measured. Coding of data***
Dependent		
	Harmhelp	Q.4. (-2) – 2
Independent		
	Familiar	Q.2 (-2) – 2
	Imp	Q.5. (-2) – 2, this question has 10 different variables, one for each subquestion, Imphealth, Impeduc, Impfins, Impfina, Impretail, Imptrans, Impmanu, Implegal, Impmedent, and Impgov
	Betterdec	Q.6. (-2) – 2, this question has 5 different variables, one for each subquestion, betfin, bethealth, betlegal, betempl and betpers
	Concern	Q.7. 1 – 5
	Betterout	Q.8. (-2) – 2
	Riskbehav	Q.9. 0 – 10
	HumorAI	Q.14. 1 or 0, 1 if prefer AI over human at 50% probability of success
	Female	Q.17. 1 if female
	Education	Q.20. 1 – 6 for education, 0 for do not prefer to say and other
	Income	Q.21. 1 – 5, 0 for do not prefer to say

Note: Each alternative in Q.5 are abbreviations for each subquestion. Question measured refers to in which question the data was captured and Coding of data refers to how the data was recoded.

Table 4.9 displays a regression model with all available explanatory variables and a constant that serves as an intercept. The model identifies eight (8) significant variables and an r-squared of 0.733. Among these variables, five (5) of these variables are affecting in a positive way, while three (3) have a negative effect. Among the significant variables, 'Betterout' has the highest positive coefficient, while 'Concern' has the highest negative coefficient.

Table 4.9 Regression, all variables

Regression, all variables	***Coefficient***	***Std.error***	***T-ratio***	***P-value***
Dependent variable				
Harmhelp				
Independent variables				
Constant	0.439	0.340	1.289	0.201
Familiar	-0.002	0.067	-0.024	0.981
Imphealth	0.015	0.083	0.176	0.861
Impeduc	0.144	0.053	2.691	0.009(**)
Impfins	0.029	0.089	0.329	0.743
Impfina	0.189	0.089	2.099	0.039(*)
Impretail	0.082	0.069	1.199	0.234
Imptrans	0.020	0.102	0.197	0.844
Impmanu	0.143	0.089	1.597	0.114
Implegal	0.088	0.081	1.086	0.280
Impmedent	0.149	0.060	2.480	0.015(*)
Impgov	-0.151	0.072	-2.101	0.038(*)
Betfin	0.064	0.083	0.774	0.441
Bethealth	-0.008	0.079	-0.103	0.918
Betlegal	-0.084	0.074	-1.142	0.257
Betempl	-0.064	0.066	-0.966	0.337
Betpers	0.038	0.072	0.521	0.604
Concern	-0.169	0.062	-2.715	0.008(**)
Betterout	0.199	0.089	2.247	0.027(*)
Riskbehav	0.069	0.028	2.443	0.017(*)
HumorAI	-0.054	0.124	-0.441	0.660
Female	0.079	0.139	0.571	0.569
Education	-0.022	0.044	-0.497	0.621
Income	-0.111	0.041	-2.729	0.008(**)
Observations	114			
R-squared	0.733			

Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001.

Figure 4.10 shows the correlation matrix of the variables, high correlation is noticed between Impfina – Impeduc, Impmanu – Imptrans, Betlegal – Implegal, and Impgov – Implegal. These variable pairs have correlation coefficients ranging from 0.65 to 0.74, this is expected due to the similarity of the questions.

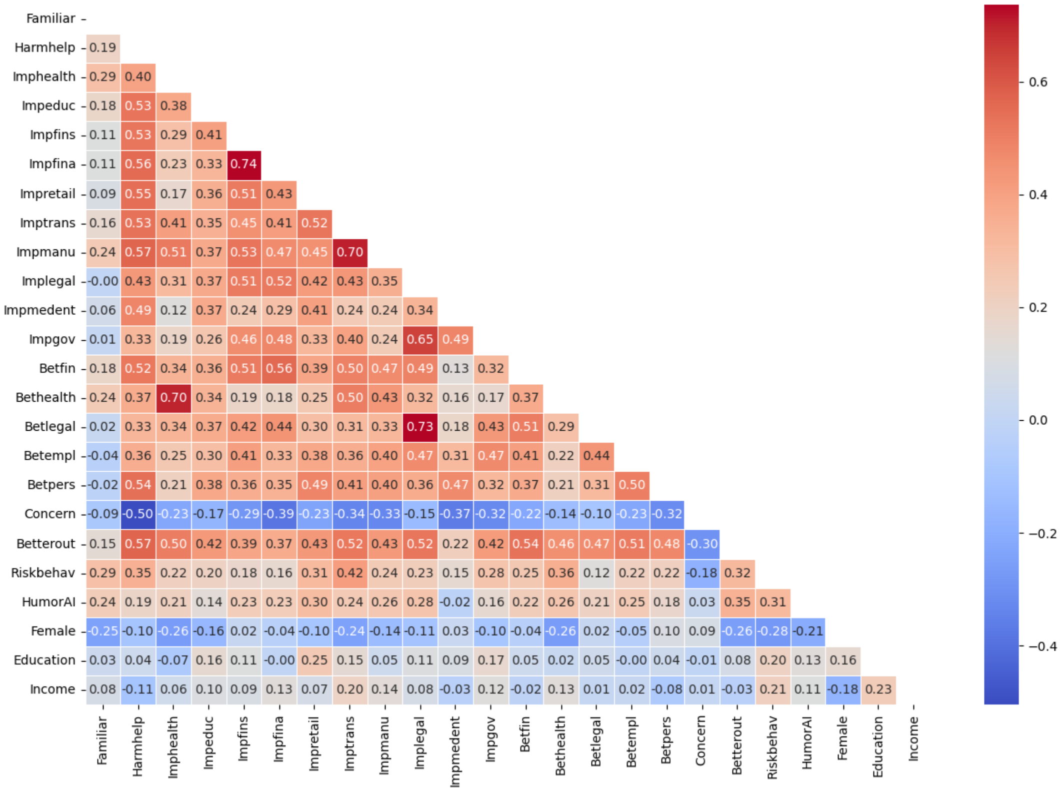

Figure 4.10 Correlation matrix

4.3.1 Backward Selection

To identify the most important variables the method of backward selection was used. This method starts with the initial regression which has all possible variables and eliminates variables one at a time based on their statistical significance. This procedure is repeated until only the most important variables are left in the model. Methods such as this one is used to simplify and enhance the interpretation of a model. Equation 4.2 presents the model after the backward selection has been performed.

$$\begin{aligned} Harmhelp = {} & Constant + \beta_1 Impeduc + \beta_2 Impfina + \beta_3 Impmanu + \beta_4 Impmedent \\ & + \beta_5 Impgov + \beta_6 Concern + \beta_7 Betterout + \beta_8 Riskbehav + \beta_9 Income \\ & + \varepsilon \end{aligned}$$

Equation 4.2

Table 4.10 provides a more detailed presentation of the model, which includes a constant and nine (9) variables. All variables are statistically significant, with either a p-value of 0.05 or less, or a p-value of 0.001 or less. The initial model, which included 23 variables and a constant, had an r-squared value of 0.733. The simplified model, with only nine variables and a constant, has a slightly lower r-squared of 0.713. Notably, three variables, Impgov, Concern, and Income have a negative effect, while Impeduc, Impfina, Impmanu, Impmedent, Betterout, and Riskbehav have positive effects. The variable with the highest coefficient excluding the intercept, is Impfina, while the variable with the highest negative effect is Concern. This represents a change from the initial model, where Betterout had the highest positive coefficient.

Table 4.10 OLS, backward selection

OLS, backward selection	***Coefficient***	***Std.error***	***T-ratio***	***P-value***	***VIF-factor***
Dependent variable					
Harmhelp					
Independent variables					
Constant	0.370	0.239	1.551	0.124	
Impeduc	0.146	0.048	3.067	0.003(**)	1.459
Impfina	0.257	0.063	4.064	0.000(***)	1.667
Impmanu	0.178	0.062	2.884	0.005(**)	1.544
Impmedent	0.181	0.050	3.637	0.000(***)	1.578
Impgov	-0.151	0.056	-2.712	0.008(**)	1.800
Concern	-0.160	0.056	-2.864	0.005(**)	1.362
Betterout	0.217	0.068	3.194	0.002(**)	1.660
Riskbehav	0.074	0.024	3.062	0.003(**)	1.221
Income	-0.123	0.037	-3.336	0.001(***)	1.136
Observations	114				
R-squared	0.713				

Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001.

The levels of the VIF factor and the correlation matrix indicate that there is low multicollinearity in the model. However, the Breusch-Pagan test for heteroscedasticity yields a significant result at the 0.05 significance level with a p-value of 0.0034, indicating the presence of heteroscedasticity in the data. Therefore, robust standard errors are used, which are presented in Table 4.11. All variables remain statistically significant.

Table 4.11 OLS, Robust standard errors

OLS, robust standard errors	***Coefficient***	***Std.error***	***T-ratio***	***P-value***
Dependent variable				
Harmhelp				
Independent variables				
Constant	0.370	0.307	1.204	0.231
Impeduc	0.146	0.045	3.263	0.002(**)
Impfina	0.257	0.068	3.802	0.000(***)
Impmanu	0.178	0.086	2.069	0.041(*)
Impmedent	0.181	0.048	3.803	0.000(***)
Impgov	-0.151	0.055	-2.725	0.008(**)
Concern	-0.160	0.059	-2.722	0.008(**)
Betterout	0.217	0.065	3.337	0.001(**)
Riskbehav	0.074	0.031	2.383	0.019(*)
Income	-0.123	0.032	-3.877	0.000(***)
Observations	114			
R-squared	0.713			

Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001.

Figure 4.11 shows a plot of the residuals, which reveals the presence of two outliers. One outlier appears in the upper right corner of the plot, while the other is located in the lowest right corner.

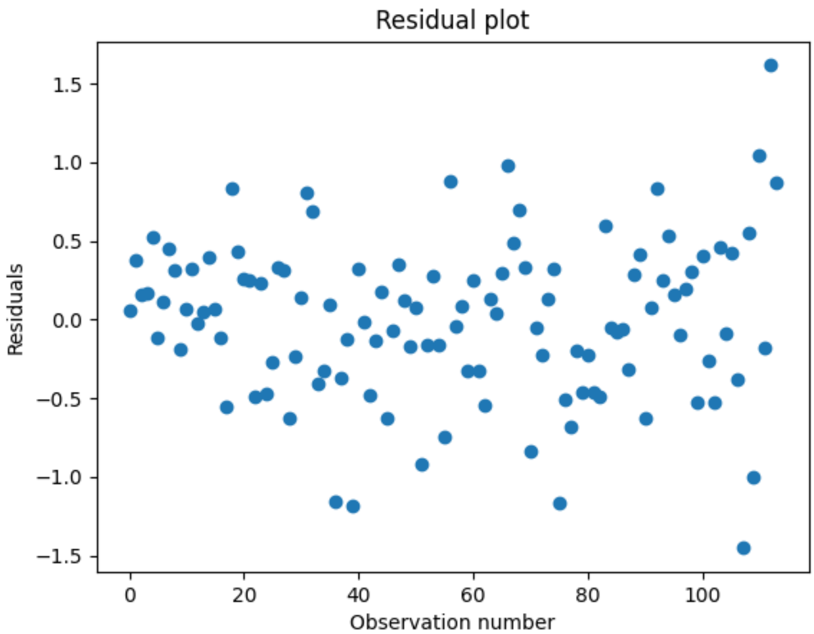

Figure 4.11 Residuals

After removing the outliers, the backward selection method was re-run to obtain the final model, which is specified in equation 4.3. Detailed information about the model is provided in Table 4.12.

$$Harmhelp = Constant + \beta_1 Impeduc + \beta_2 Impfins + \beta_3 Impfina + \beta_4 Imptrans + \beta_5 Impmedent + \beta_6 Impgov + \beta_7 Concern + \beta_8 Betterout + \beta_9 Riskbehav + \beta_{10} Income + \varepsilon$$

Equation 4.3

Table 4.12 OLS, final

OLS, final	***Coefficient***	***Std.error***	***T-ratio***	***P-value***	***95% C.I.***	***VIF-factor***
Dependent variable Harmhelp						
Independent variables						
Constant	0.030	0.227	0.133	0.894	(-0.419) - 0.479	
Impeduc	0.103	0.045	2.290	0.024(*)	0.014 – 0.192	1.532
Impfins	0.144	0.071	2.035	0.044(*)	0.004 – 0.285	2.572
Impfina	0.221	0.070	3.159	0.002(**)	0.082 – 0.359	2.475
Imptrans	0.146	0.068	2.132	0.035(*)	0.010 – 0.282	1.805
Impmedent	0.210	0.045	4.630	0.000(***)	0.120 – 0.299	1.586
Impgov	-0.224	0.052	-4.320	0.000(***)	(-0.327) – (-0.121)	1.893
Concern	-0.118	0.052	-2.269	0.025(*)	(-0.221) – (-0.015)	1.375
Betterout	0.246	0.063	3.890	0.000(***)	0.121 – 0.371	1.746
Riskbehav	0.098	0.023	4.177	0.000(***)	0.051 – 0.145	1.322
Income	-0.112	0.034	-3.316	0.001(***)	(-0.179) – (-0.035)	1.149
Observations	112					
R-squared	0.755					

Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001. C.I. represents the 95% confidence interval.

The final model included some changes in variables. Impfins, and Imptrans were added while Impmanu was removed. The model was tested for heteroscedasticity using Breusch-Pagan, yielding a p-value of 0.089. This p-value indicates that the null hypothesis of heteroscedasticity not present cannot be rejected on the 0.05 significance level. Nevertheless, all variables remained statistically significant. It is noteworthy that the VIF factor for Impfins and Impfina exceeded 2, and the correlation between these variables was 0.75. However, since the VIF factor remained relatively low, neither variable was excluded from the final model. The model showed a slight increase in explanatory power with an r-squared of 0.755.

4.3.2 Alternative Model

Due to the obvious similarity between the variables and the Imp-variables, a simplified and alternative model is presented in Table 4.13. This model demonstrates consistent patterns as previous models. Except for the constant, Betterout is the variable with the largest coefficient and concern is the variable with the highest negative coefficient. The model suggests a negative relationship between Harmhelp, and the variables Concern and Income. I.e., for every unit increase in Concern and all else being equal, individuals perceive AI as more harmful. This indicates there is a relationship between the individuals that are highly concerned about AI and believing that AI is harmful. Similarly, self-assessed risk behavior is significant and therefore a positive relationship between risk behavior and belief whether AI is helpful or harmful is established. More risk-taking individuals perceive AI as more helpful than others. This model exhibits a smaller r-squared than previous models, this is expected as the backward selection method by construction selects the variables that are statistically significant and therefore, is expected to have a large r-squared. When tested for heteroscedasticity the model does not yield a significant result at the 0.05 significance level. The test yields a p-value of 0.098. Additionally, the model does not yield significant results when tested for misspecification through Ramsey's Reset test, neither squares and cubes, squares only or cubes only are significant, indicating that there are no significant specification errors in the model according to the test. In contrast to the OLS, final model: the alternative model does not yield a significant result for the variable Income, however, the relationship with the dependent variable remains negative, with a small deviation in the 95% confidence interval.

Table 4.13 OLS, alternative model

OLS, alternative model	***Coefficient***	***Std.error***	***T-ratio***	***P-value***	***95% C.I.***	***VIF-factor***
Dependent variable Harmhelp						
Independent variables						
Constant	0.541	0.361	1.501	0.136	(-0.174) – 1.256	
Familiar	0.039	0.073	0.533	0.595	(-0.106) – 0.185	1.143
Concern	-0.253	0.065	-3.925	0.000(***)	(-0.381) – (-0.125)	1.139
Betterout	0.427	0.073	5.824	0.000(***)	0.281 – 0.572	1.258
Riskbehav	0.117	0.033	3.583	0.000(***)	0.052 – 0.182	1.381
Female	0.160	0.149	1.081	0.282	(-0.134) – 0.455	1.247
Education	-0.049	0.049	-1.000	0.320	(-0.145) – 0.048	1.203
Income	-0.073	0.046	-1.586	0.116	(-0.165) – 0.018	1.153
Observations	112					
R-squared	0.529					

Significance levels are denoted by asterisks (*). One asterisk (*) indicates a p-value less than 0.05, two asterisks (**) indicate a p-value less than 0.01, and three asterisks (***) indicate a p-value less than 0.001. C.I. represents the 95% confidence interval.

5 Discussion

This experimental study aims to analyze individuals' perceptions of AI using concepts from behavioral economics related to decision-making. Statistical models, such as regression analysis, were used to analyze data collected from an experimental study with participants recruited from an online panel obtain through Prolific. This study aims to answer the following research question:

'Are individuals more likely to perceive potential risks associated with artificial intelligence than potential benefits?'

The experimental survey found that individuals are to some extent familiar with AI. This is coherent with the increasing amount of publications and publicity AI has received in recent years. Moreover, a considerable amount of the participants had used an AI service before.

The final OLS regression, which excluded two outliers and selected the most important variables, showed that seven variables had a significant positive impact on individuals' perception of AI as helpful, these were Impeduc, Impfins, Impfina, Imptrans, Impmedent, Betterout, and Riskbehav. The first five variables indicated that individuals believed that AI could have a positive impact and be helpful. Betterout and Riskbehav also showed a positive relationship, with the latter variable capturing individuals' self-assessed risk behavior. Individuals who were more willing to take risks were more likely to perceive AI as helpful. On the other hand, the variables Impgov, Concern, and Income had a significant negative impact on individuals' perception of AI. Individuals were not optimistic about AI's impact on the government sector, and there was a significant relationship between concern about AI and belief in whether it was harmful.

The study suggests that individuals, in general, had a positive view of AI and perceived it as useful, with the regression providing a strong r-squared, indicating good explanatory power in the model

The alternative OLS model reveals similar relationships to the OLS final model. However, there is one notable distinction between the two models: the variable Income is found to be statistically insignificant. It is important to note that although its insignificance, the variable exhibits just a minor variation in the confidence interval. Thus, it should not be completely disregarded as an important variable.

Questions 12 and 13 of the experiment framed the participants in a positive respectively a negative way. The difference of mean analysis suggests that the negative framing had a considerable effect on the respondents, from a behavioral economics perspective this can be explained by framing and a shift of reference point. As explained by Kahneman and Tversky (1986) framing and reference points are important in human decision-making. The negative framing observed in question 13 could indicate that participants did not previously consider AI's potential negative impacts, or that they felt a social norm of having a more negative view of AI. Notably, as the positive and negative framing are not complete opposites, one can argue that the positive framing may be the object of a more subjective interpretation.

Question 14 revealed that there is no clear preference for an AI advisor or a human advisor. The inflection point is where both advisors have a 50% success probability, below 50% success probability the individuals prefer the AI advisor, while above 50% they prefer the human advisor. At exactly 50% success probability, the data indicate no obvious preference, there is a slight preference for the human advisor, as 54.4% of the participants chose the human advisor, while the AI advisor was selected by 45.6% of the participants. The positive attitude towards AI can be explained by the significant variables Impfins and Impfina in the OLS regression, which had a positive impact on the perception of AI.

Figure 4.8 and Figure 4.9 illustrate that the two treatments applied to question 14 did not produce the expected result. In the negative treatment group, respondents tended to choose AI as the investor at a 50% success probability, whereas the positive treatment group leaned towards the human investor. However, no significant difference was found between these two treatment groups. According to Kahneman and Tversky (1986), factors such as language, context, and display can influence the analysis of framing effects, which might have influenced the outcomes. It is important to consider that the treatment may not have been optimal, or the readers might not have given sufficient thought to the subsequent questions in the survey. However, as the human investor's success probability decrease below 50%, for example at 47.5%, there is an obvious difference between the two treatment groups. At 47.5% the negative treatment group prefers the human investor to a larger extent than the positive framing group and vice versa. Even though choosing the human investor is an objectively worse option. Thus, the experiment shows weak signs of successful framing. However, as the effectiveness of the treatments is not significant, further analysis is needed to explore whether individuals' perceptions of AI are robust. Such analysis could include other framings or longitudinal studies.

In terms of risk aversion, a majority of the total sample displayed risk-averse behavior. However, when comparing the two treatment groups, a slight difference is observed. In the positive treatment group, 59% of participants exhibited risk-averse behavior, while in the negative treatment group, 51% displayed risk-averse behavior. Despite these minor variations, the impact of participants' risk aversion concerning question 14 remains inconclusive.

In conclusion, this study can offer insights into individuals' perceptions of AI. As the interest in AI grows in society, policymakers and businesses can use this information to understand areas where a positive or negative attitude toward AI is expected. However, the concerns raised in the open letter written and signed by Elon Musk and others should be taken into consideration, as these concerns are voiced by business leaders and researchers within the field of AI who perhaps has great knowledge of the future developments of AI and its capabilities.

5.1 Limitations

To overcome limitations in the survey experiment, a pilot study was conducted to assess the questions' interpretability and make necessary adjustments. The pilot revealed one obvious issue, the attention check. Therefore, the attention check was revised and as a result, only 5% of the participants failed the attention check in the final experiment. This was a significant improvement from the pilot study.

In addition to the limitation discussed, other factors may have affected the analysis and findings. One such factor is the sample size, which may have been too small to support generalizations of the results to a broader population. Nevertheless, statistical significance was achieved for multiple variables. Another limitation is the subjective interpretation of answer options, which could introduce bias or inconsistency in how participants responded to survey questions. These limitations should be considered when interpreting the results of this study.

6 Conclusion

This study aimed to examine perceptions of risks and potential benefits associated with AI. Adopting a perspective from behavioral economics, an experimental study was conducted to analyze these perceptions. The study employed two treatment effects to examine the impact of positive and negative framing of AI. The participants were obtained from Prolific and incentivized with a smaller payment. A total of 120 respondents completed the study, and the analysis focused on 114 of these participants.

The results indicate a negative relationship between the dependent variable, which measures individuals' belief about whether AI is helpful or harmful, and the independent variables, income, and how concerned individuals are about AI. On the other hand, a positive relationship is found with the independent variables that measure self-assessed risk behavior, and beliefs regarding whether AI will create better outcomes or not. Furthermore, while self-assessed risk behavior is found to be a significant predictor, no gender difference is observed. Additionally, the difference of mean results reveals a significant difference in beliefs about whether AI is harmful or helpful. However, no significant framing effect is found in the subsequent survey questions, only minor differences are observed. Furthermore, there is no significant difference in risk aversion between the two treatment groups.

To further investigate the perception of AI in society, future research could explore gender differences by utilizing a more comprehensive analysis of risk behavior. Additionally, future research could focus on standardizing and evaluating how the perception and other potential disruptors evolve.

7 Bibliography

Dash, Y., Mishra, S.K., and Panigrahi, B.K. (2018). Rainfall prediction for the Kerala State of India using artificial intelligence approaches. *Computers & Electrical Engineering,* 70, pp. 66-73. Available online: https://doi.org/10.1016/j.compeleceng.2018.06.004.

Ding, X., Hartog, J., and Sun, Y. (2010). Can We Measure Individual Risk Attitudes in a Survey?' *SSRN Electronic Journal.* Available online: https://doi.org/10.2139/ssrn.1570425.

Dohmen, T., Armin, F., Huffman, D., Sunde, U., Schupp, J., and Wagner, G.G. (2005). Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. *SSRN Electronic Journal*. Available online: https://doi.org/10.2139/ssrn.807408.

Došilović, F.K., Brčić, M. and Hlupić, N. (2018). Explainable artificial intelligence: A Survey. *2018 41st International Convention of Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Available online: https://doi.org/10.23919/mipro.2018.8400040.

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. Available online: https://futureoflife.org/open-letter/pause-giant-ai-experiments/. [Accessed: 2023-04-26].

Hartman, R.S., Doane, M.J., and Woo, C.-K. (1991). Consumer rationality and the status quo, *The Quarterly Journal of Economics,* 106(1), pp. 141-162. Available online: https://doi.org/10.2307/2937910.

Haaland, I., Roth, C., and Wohlfart. J. (2023). Designing Information Provision Experiments. *Journal of Economic Literature*, 61(1), pp. 3-40. Available online: https://doi.org/10.1257/jel.20211658.

Hu, K. (2023). ChatGPT sets record for fastest-growing user base – analyst note. *Reuters*, 2 February. Available online: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/. [Accessed: 2023-04-22].

Institute of Electrical and Electronics Engineers. (2019). IEEE Position Statement Artificial Intelligence. Available online: https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf

Jolls, C., Sunstein, C.R., and Thaler, R.H. (2000). A Behavioral Approach to Law and Economics. *Behavioral Law and Economics*, pp.13-58. Available online: https://doi.org/10.1017/cbo9781139175197.002.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review,* 93(5), pp. 1449-1475. Available online: https://doi.org/10.1257/000282803322655392.

Kahneman, D., Knetsch, Jack, L., and Thaler, H, R. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives,* 5(1), pp. 193-206. Available online: https://doi.org/10.1257/jep.5.1.193.

Kahneman, D. and Tversky, A. (1974). Judgment under uncertainty: Heuristics and Biases. *Science,* 185(4157), pp. 1124-1131. Available online: https://doi.org/10.1126/science.185.4157.1124.

Kahneman, D. and Tversky, A. (1979). Prospect theory, an analysis of decision under risk. *Econometrica*, 47(2), pp. 264-291.

Kahneman, D. and Tversky, A. (1986). Rational choice and the framing of decisions. *The Journal of Business,* 59(4), pp. 251-278. Available online: https://doi.org/10.1086/296365.

Klaes, M., and Wilkinson, N. (2018). An Introduction to Behavioral Economics. 3rd edition. Palgrave.

Knetsch, J.L. (1989). The Endowment Effect and Evidence of Nonreversible Indifference Curves. *The American Economic Review,* 79(5), pp. 1277-1284. Available online: http://www.jstor.org/stable/1831454.

Krosnick, A.J. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), pp. 213-236.

Kurzweil, R. (1985). What is Artificial Intelligence Anyway? As the techniques of computing grow more sophisticated, machines are beginning to appear intelligent – but can they actually think?. *American Scientist*, 73(3), pp. 258-264. Available online: https://jstor.org/stable/27853237.

Liu, C., Liu, X., Wu, F., Xie, M., Feng, Y., and Hu, C. (2018). Using Artificial Intelligence (Watson For Oncology) for Treatment Recommendations Amongst Chinese Patients with Lung Cancer: Feasibility Study. *Journal of Medical Internet Research*, 20(9). Available online: https://doi.org/10.2196/11087.

Lu, Y. (2019). Artificial Intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), pp. 1-29. Available online: https://doi.org/10.1080/23270012.2019.1570365.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Carlos Niebles, J., Parli, V., Shoham, Yoav., Wald, R., Clark, J., and Perrault,

R. (2023). The AI Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI. Stanford University.

Mata, J., Miguel, d.I., Durán, J. R., Merayo, N., Singh, K.S., Jukan, A., and Chamania, M. (2018). Artificial intelligence (AI) methods in optical networks: A comprehensive study. *Optical Switching and Networking,* 28, pp. 43-57. Available online: https://doi.org/10.1016/j.osn.2017.12.006.

Posner, R.A. (1998). Rational choice, behavioral economics, and the law. *Stanford Law Review,* 50(5), p. 1551-1575. Available online: https://doi.org/10.2307/1229305.

Prolific. (2022). About Us. Available online: https://www.prolific.co/about. [Accessed: 2023-04-27].

Samuelson, W., and Zeckhauser, R. (1988). Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty,* 1(1), pp. 7-59. Available online: https://doi.org/10.1007/bf00055564.

Sharbek, N. (2022). How Traditional Financial Institutions have adapted to Artificial Intelligence, Machine Learning and FinTech?. *Proceedings of the international Conference on Business Excellence*, 16(1), pp. 837-848. Available online: https://doi.org/10.2478/picbe-2022-0078.

Simon, H. A. (1957). Models of man; social and rational: mathematical essays on rational human behavior in a social setting. New York: John Wiley & Sons.

Smith, D. (2002). Psychologist wins Nobel Prize. *American Psychological Association.* Available online: https://www.apa.org/monitor/dec02/nobel.html. [Accessed: 2023-04-28].

Stanovich, K.E., and West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences*, 23(5), pp. 645-665. Available online: https://doi.org/10.1017/s0140525x00003435.

Steele, C.M., and Aronson, J. (1995). Stereotype threat and intellectual test performance of African Americans. Journal of personality and social psychology, 69(5), pp. 797-811.

Stefánsson, H.O., and Bradley, R. (2019). What is risk aversion?. *The British Journal for the Philosophy of Science*, 70(1), pp. 77-102. Available online: https://doi.org/10.1093/bjps/axx035.

Stilgoe, J. (2017). Machine Learning, social learning and the governance of self-driving cars. *SSRN Electronic Journal*. Available online: https://doi.org/10.2139/ssrn.2937316.

Tan, K.-H., and Lim, B.P. (2018). The artificial intelligence renaissance: deep learning and the road to human-Level machine intelligence. *APSIPA Transactions on Signal and Information Processing,* 7(1). Available online: https://doi.org/10.1017/atsip.2018.6.

Appendix A: Survey

Perceptions of artificial intelligence

Thank you for taking the time to participate in this study. I would like to assure you that your responses will remain anonymous and confidential. The purpose of this survey is to gather information for a Master's Thesis project within artificial intelligence (AI) and financial services at Lund University, School of Economics and Management.

The collected data will be used solely for research purposes and only the researcher will have access to it, the data will be presented on group level. Participation in this survey is voluntary. By submitting the completed survey, you are providing consent for the researcher to use your responses for research purposes.

The survey aims to gather information about individuals' beliefs, familiarity, and concerns regarding artificial intelligence. The survey consists of 19 questions and is expected to take around 8 minutes to complete.

If you have any questions about the survey or the research project, please feel free to contact the researcher via email at vi2845kr-s@student.lu.se. Thank you again for your participation in this study.

Section 1/5

1. What is your unique Prolific ID?

____

2. How familiar are you with artificial intelligence?

   1 - Not at all familiar
   2 - Somewhat familiar
   3 - Moderately familiar
   4 - Very familiar
   5 - Extremely familiar

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Extremely familiar

3. Which of the following AI-powered services have you used in the recent year? Note: You can check multiple options.

- [ ] ChatGPT
- [ ] Bard AI
- [ ] DALL-E
- [ ] AlphaCode
- [ ] Gato
- [ ] AlphaTensor
- [ ] LaMDA
- [ ] CoPilot
- [ ] Bing AI
- [ ] ChatSonic
- [ ] Jasper Chat
- [ ] Socratic
- [ ] DialoGPT
- [ ] I have not used any AI service or chatbot in the recent year
- [ ] Other: ____________________

4. Do you think artificial intelligence is mostly helpful or mostly harmful to people?

1 - Mostly harmful
2 - Somewhat harmful
3 - Neither helpful nor harmful
4 - Somewhat helpful
5 - Mostly helpful

	1	2	3	4	5	
Mos	◯	◯	◯	◯	◯	Mostly helpful

5. In the sectors specified below, do you think artificial intelligence will have a positive or negative impact?

	Mostly	Somewhat	Neither positive	Somewhat	Mostly

	negative	negative	nor negative	positive	positive
Healthcare (e.g. medical services, hospital and related industries)					
Education (e.g. schools, universities and other educational institutions)					
Financial services (e.g. banking, insurance and other financial institutions)					
Financial advice (e.g. financial planning, investment management and other financial advisory services)					
Retail (e.g. sale of goods and services)					
Transportation (e.g. shipping, logistics and public transportation)					
Manufacturing (e.g. production of machinery and industrial processes)					
Legal (e.g. law firms and legal services)					
Media and entertainment (e.g. television, film, music and other forms of entertainment)					
Government (e.g. public administration, governance and policy-making)					

6. To what extent do you agree or disagree with 'Artificial intelligence will help people make better decisions in the following contexts?'

	Strongly disagree	Disagree	Neither agree nor disagreee	Agree	Strongl agree
Financial decisions (e.g. budgeting, investments, loans)					
Healthcare decisions (e.g. medical diagnoses, treatment plans)					
Legal decisions (e.g. contracts, litigation)					
Employment decisions (e.g. hiring, performance evaluations)					
Personal life decisions (e.g. travel plans, purchasing decisions)					

7. How concerned are you about the use of artificial intelligence? *

1 - Not concerned at all
2 - Slightly concerned
3 - Somewhat concerned
4 - Quite concerned
5 - Very concerned

	1	2	3	4	5	
Not						Very concerned

8. To what extent do you agree or disagree with 'it is possible for artificial intelligence to help humans achieve better outcomes than human beings working alone'?

1 - Strongly disagree
2 - Somewhat disagree
3 - Neither agree nor disagree
4 - Somewhat agree
5 - Strongly agree

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

9. How do you see yourself: Are you in general a person who takes risk or do you try to evade risks?

0 - Unwilling to take risks
10 - Fully prepared to take risks

	0	1	2	3	4	5	6	7	8	9	10	
Unw	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Fully prepared to take risks

Section 2/5

Please read this section carefully as you will be asked two questions and one answer will be provided in the text.

The first question is about a problem. In questionnaires, sometimes there are participants who do not carefully read the questions, this could compromise the study. To show that you read the questions carefully, please enter number five (5) as the answer to the next question, what is your favorite number?

The second question will ask you to state if you are born in an 'odd' or 'even' month.

10. What is your favorite number?

- 1
- 2
- 3
- 4
- 5

11. Were you born in an 'odd' or 'even' month?

Odd: January, March, May, July, September, and November.
Even: February, April, June, August, October, and December.

- Odd
- Even

Section 3/5

12. In a report by Maslej et al. (2023) the main reasons Americans are excited about artificial intelligence are the following: artificial intelligence will make life and society better, help humans being more efficient, help with difficult tasks and artificial intelligence is more accurate than humans.

    Do you think artificial intelligence is mostly helpful or mostly harmful?

    1 - Mostly harmful
    2 - Somewhat harmful
    3 - Neither helpful nor harmful
    4 - Somewhat helpful
    5 - Mostly helpful

	1	2	3	4	5	
Mos	○	○	○	○	○	Mostly helpful

Section 3/5

13. In a report by Maslej et al. (2023) the main reasons Americans are concerned about artificial intelligence are the following: artificial intelligence will create loss of human jobs, enable surveillance and hacking, and reduce digital privacy. Artificial intelligence will also be misused and become too powerful.

    Do you think artificial intelligence is mostly helpful or mostly harmful?

    1 - Mostly harmful
    2 - Somewhat harmful
    3 - Neither helpful nor harmful
    4 - Somewhat helpful
    5 - Mostly helpful

	1	2	3	4	5	
Mos	○	○	○	○	○	Mostly helpful

Section 4/5

In this section you are asked to answer a two hypothetical questions. Imagine that you are faced wit the actual problem and indicate the decision you would have made in that case. There are no correct answers in these problems and I want to remind you that your answers remain anonymous.

14. Imagine you had won 1000 EUR in a lottery. Immediately after, a reputable bank is giving you the opportunity to invest and potentially double your investment in one year.
The bank asks whether you would prefer your investment to be done by a human advisor or by an AI-advisor. Based on past experience, the bank estimates that the AI-advisor has a 50% chance of a successful investment. Human advisors' performance can vary more, with some performing better and some worse than the AI. In case the investment is unsuccessful, you will lose the entire investment.

    Therefore, the bank asks "how successful must the human advisor be for you to prefer the human advisor over the AI?"

    Please indicate which advisor you prefer given the probabilities below.

	Human Advisor	AI-Advisor (50%)
Human 60%	◯	◯
Human 57.5%	◯	◯
Human 55%	◯	◯
Human 52.5%	◯	◯
Human 50%	◯	◯
Human 47.5%	◯	◯
Human 45%	◯	◯
Human 42.5%	◯	◯
Human 40%	◯	◯

15. Would you like to participate in a lottery where you have 50% chance to win 10 EUR and 50% chance to loose 10 EUR?

- Yes
- No

Section 5/5

In this section you will be asked to answer a couple of demographic questions.

16. How old are you?

- 0-10
- 11-20
- 21-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71+
- Prefer not to say

17. What is your sex?

- Male
- Female
- Non-binary
- Prefer not to say
- Other

18. What is your occupation?

- Student
- Employed
- Entrepreneur
- Seeking employment
- Unemployed
- Retired
- Prefer not to say
- Other

19. What sector do you work in or operate your business in?

- Financial operations, business services
- Human health and social work
- Education
- Manufacturing, mining and quarrying, energy and environment
- Trade
- Public administration
- Information and communication
- Personal and cultural services
- Transport
- Accommodation and food services
- Agriculture, forestry, fishing
- Other

20. What is the highest level of education you have completed?

- Lower secondary education (e.g., middle school)
- Upper secondary education (e.g., high school)
- Vocational education (e.g., trade school)
- Bachelor's degree (e.g., university undergraduate degree)
- Master's degree (e.g., university graduate degree)
- Doctoral degree (e.g., PhD)
- Other
- Prefer not to say

21. What is your annual income?

- Less than 10 000 EUR
- Between 10 001 EUR and 30 000 EUR
- Between 30 001 EUR and 50 000 EUR
- Between 50 001 EUR and 70 000 EUR
- More than 70 001 EUR
- Prefer not to say