Universiteit Leiden
The Netherlands

Computer Science & Economics

Guidelines for designing experiments to study Human-AI interactions

Ruben Visser

Supervisors:
Amirhossein Zohrehvand & Niki van Stein

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

23/07/2024

Abstract

This thesis establishes guidelines for designing experiments to study interactions between humans and AI agents. As AI becomes increasingly integrated into modern society, understanding these interactions is essential for improving the effectiveness and accuracy of Human-AI interactions. Despite the growing number of studies in this area, clear guidelines for the experimental design are lacking. This can affect the reliability and validity of the experimental results. To address this issue, the thesis conducts an extensive literature review on Human-AI interactions. We focus on the relation between humans and AI agents and the tasks that define the interactions. By distinguishing between coordination, cooperation, and collaborative interactions, and examining task dimensions such as speed, complexity, and importance, this thesis proposes detailed guidelines for experimental design of Human-AI interactions. These guidelines cover the experimental approach and the requirements for successful interaction for the different interaction types and task dimensions including which model to use for the AI agent and what requirements the platform for the interaction experiment must meet. The findings offer a framework that researchers can use to achieve more reliable and valid experimental results in studying Human-AI interactions.

Contents

1 Introduction

The use of experiments as a research method has become increasingly prevalent in science (Webster & Sell, 2014). In recent years, various studies have examined how experiments should be designed (Levine et al., 2023; Berger et al., 2018) to ensure that they yield reliable and valid results. These studies have focused on establishing general guidelines for properly designing and setting up various types of experiments. Additionally, specialised guidelines have been developed for specific scientific fields (Grootswagers, 2020; Klasnja et al., 2015; Kohavi et al., 2020) addressing the unique requirements of each field.

However, the experimentation surrounding interactions between humans and artificial intelligence (AI) agents[1], i.e. Human-AI interactions, remain underdeveloped in terms of standardised guidelines. Despite the increasing number of experiments conducted in this field (Ashktorab et al., 2020; Carroll et al., 2020; Fügener et al., 2021; Gnewuch et al., 2023; Han et al., 2022; Li et al., 2022; J. Liu et al., 2024; Lucas et al., 2014), there is a noticeable gap in clear experimental guidelines tailored specifically to Human-AI interactions. It is important to have specific guidelines for this field in addition to general guidelines due to the numerous nuances involved in these interactions, as the reliability and validity of experimental results are at stake when an experiment is improperly designed (Tanco et al., 2009). Every interaction is different and has specific characteristics that must be carefully considered. Addressing this gap is crucial for advancing our understanding of how humans interact with AI agents and for ensuring that experimental results are both reliable and comparable across different studies. This leads to the research question: *what are the guidelines for designing experiments to study Human-AI interactions?*

The purpose of this thesis is therefore to provide guidelines for designing Human-AI interaction experiments. Previous research has conceptualised these interactions based on the relation between humans and AI agents. Using this perspective, three different types of Human-AI interactions are distinguished: coordination, collaboration and cooperation (Schmidt & Loidolt, 2023). This perspective only shows how the relation between humans and AI agents influences the way they interact per interaction type. We therefore suggest in this thesis to also consider the task around which the interaction revolves. The task influences how an interaction unfolds, impacting factors such as interaction speed. We can distinguish tasks by looking at three

[1] An AI agent can be defined as an entity that performs actions within a specific environment (Poole & Mackworth, 2010) controlled by AI.

dimensions: the speed, complexity, and importance of the task. Together, these perspectives provide a better understanding of what these interactions entail. Using these perspectives, we are able to provide guidelines for designing experiments to study Human-AI interactions. This consists of the experimental approach and the requirements for successful interaction for both perspectives. We also consider the importance of using a suitable AI model for an AI agent that aligns with the interaction task, as well as determining the most suitable platform for these interaction experiments.

2 Human-AI Interactions

2.1 Importance

As the use of AI continues to increase in modern society, it becomes increasingly important to understand how it interacts with humans. The increase in use is reflected in economic and labour changes. For example, AI is expected to add around $15.7 trillion to the United States economy by 2030, potentially leading to a Gross Domestic Product (GDP) boost of as much as 26% (Wang et al., 2023). While in 2017 only 21% of organisations used AI in at least one business unit from a group of respondents, this has grown considerably, with a peak of 58% in 2019, to 50% in 2022 (McKinsey, 2022). Due to this increase in the use of AI in modern society, more interactions are taking place between humans and AI agents which often lead to economic benefits. Research by Choudhary et al. (2023) highlights that the accuracy of decisions derived from predictions are more precise when combining decisions made by both human and AI agents on the same predictive task, rather than solely relying on the decision of one of them. The increasing accuracy in the decisions often results in economical improvements. However, while the quality of the work increases, the originality decreases (Doshi & Hauser, 2024). In addition, although humans interacting with AI show improvement compared to those not interacting with AI, their performance declines when the task falls outside the scope of the AI (Dell'Acqua et al., 2023b). Despite these insights, there remains a significant gap in our understanding of human-AI interactions. Hence it is important to understand more about the interactions between humans and AI agents.

2.2 Conceptualisation

Considering that there are multiple parties in an interaction, past research has naturally conceptualised the Human-AI interaction based on the different forms of the relation between these two parties. Hornbæk and Oulasvirta (2017, p. 5049) define interactions as ”two entities that determine each other's behavior over time”. In Human-AI interactions these entities are AI agents and humans. This means that the actions of the human influence the actions of the AI agent, and vice versa. These interactions between humans and AI agents can take many different forms, which means that the set up and implementation of the interactions will not always be the same.

Description of Human-AI interactions is often unclear, as indicated by numerous interchangeably used terms, which are not clearly defined (Schmidt & Loidolt, 2023). One attempt to address this shortcoming is the study of Schmidt and Loidolt (2023) in which they specify various interaction types, considering that an interaction revolves around involvement of more than one party. According to this consideration, the different Human-AI interaction types include coordination, collaboration, and cooperation.

These three types differ in the achievement of objectives and execution due to differences in relation between the human and the AI agent. In coordination interactions, both the human and the AI agent pursue distinct objectives while considering each other's actions in their own decision-making processes. It is crucial to note that humans can determine or anticipate the objective of the AI agent. Collaboration interactions entail a shared objective between the human and the AI agent, established by humans. Humans give instructions to the AI agent how it should perform. Finally, in cooperative interactions, both the human and the AI agent share the same objective, which is established by humans. Nonetheless humans do not give instructions to the AI agent how it should perform, it can decide by itself (Schmidt & Loidolt, 2023).

However, another approach to conceptualising the Human-AI interaction is to take into consideration the task that these interactions intend to accomplish. Angus and Christine (2010, p. 1821) define a task as “a piece of work to be done or undertaken”. In the context of Human-AI interactions, this means that these interactions revolve around achieving and completing a specific activity or assignment. In contrast to the perspective of the parties involved, which focuses on understanding the features on how interactions occur between humans and AI agents based on the relation, the task-oriented perspective focuses on understanding the features of tasks the interactions intend to accomplish. For example, in a word guessing game where a human and an AI agent interact to guess as many words as possible, focusing solely on the perspective of the parties involved identifies it as a cooperative interaction. This perspective helps us comprehend the relation between the human and AI agent and understand how they interact. However, this approach only reveals the type of interaction occurring between humans and AI agents based on their relation. By using the task-oriented perspective, we can observe how an interaction occurs based on the specific task they aim to accomplish. This perspective allows us to gain a more comprehensive understanding of an interaction, because it presents a more detailed view of how interactions occur. This view is more detailed as it could for example illustrate how the speed of the interaction varies depending on the nature of the task.

We borrow from the strategic decision-making literature to conceptualise the task along three different dimensions: the speed, complexity and importance of the task. These dimensions arise from the three main characteristics of strategic decisions, which are ambiguity, hierarchy and irreversibility (Zohrehvand, 2020). The three characteristics are the most important when we consider prediction problems (Zohrehvand, 2020). Given that Human-AI interactions often involve predictive elements, these characteristics can effectively be used to define various dimensions of tasks. Firstly, the speed of the tasks comes from ambiguity. Ambiguity arises when we are uncertain about the probabilities associated with different outcomes (Zohrehvand, 2020). When outcomes of several steps of the task are uncertain, this leads to ambiguity. Unclear outcomes at different steps slow down the overall process as they require clarification and adjustment. This results in slower execution of the task and thus decreasing the speed of the task. Secondly, the complexity of the tasks arises from hierarchy. Hierarchy refers to a situation where one component affects the options available to another component. In that situation the first component is more hierarchical and limits the range of options available for the other components (Zohrehvand, 2020). So when one step in executing a task has a significant impact on the available options for another step of executing the task, the execution of the task becomes more hierarchical. It then becomes more complex to accomplish the task, as the task execution becomes more interdependent and less flexible. Thirdly, the importance of tasks comes from irreversibility. Irreversibility refers to the difficulty to reverse something once it is made (Zohrehvand, 2020). When the steps of execution of the task are difficult to reverse, the task is considered important because of their potentially significant consequences.

2.3 Overview of Past Research

Table 1 provides an overview of how past research can be perceived either from the features of the interaction from multiple parties or from the dimensions of the task. This overview of the past research includes the interaction types and findings, and the speed, complexity and importance of the tasks.

Table 1

Overview of How Past Research on Human-AI Interaction Can Be Perceived Either From the Features of the Interaction From Multiple Parties or From the Dimensions of the Task

Study	Interaction type	Findings	Speed	Complexity	Importance
Ashktorab et al. (2020)	Cooperation	Humans describe AI agent partners less positive than human partners, even if game outcomes are the same.	Low	High	High
Carroll et al. (2020)	Cooperation	It is important to design an AI agent specifically for interactions with humans.	High	Low	High
Dell'Acqua et al. (2023a)	Cooperation	When a human is replaced by an AI agent in a cooperative interaction, team performance declines, there are more failures in the interaction and trust is lost. Even if the AI agent performs the task better than the human.	High	Low	Low
Fügener et al. (2021)	Cooperation	Delegation only positively influences performance when the AI agent delegates, not when the human does.	Low	High	High
Gnewuch et al. (2023)	Cooperation	Humans want to leave a good impression when they know there is human involvement in the AI agent	Moderate	Low	Low
Han et al. (2022)	Cooperation	Humans are more satisfied when human agents express positive emotions, but the effectiveness diminishes when AI agents do the same.	High	Low	High

Study	Interaction type	Findings	Speed	Complexity	Importance
Li et al. (2022)	Cooperation	Humans prefer cooperative interactions with AI agents where their actions influence each other, rather than interactions where they work independently.	Variable	High	High
J. Liu et al. (2024)	Collaboration	Combining different AI models improves the outcome of the interactions.	High	Low	High
Lucas et al. (2014)	Cooperation	Humans feel more comfortable when interacting with AI agents than with humans	High	Low	High

2.3.1 Interaction Type and Speed, Complexity and Importance of the Task

Various past studies have investigated interactions differently. These studies can be perceived differently based on the interaction type and speed, complexity and importance of the task. In the study by Ashktorab et al. (2020), a cooperative interaction takes place in a digital word prediction game where both the human and AI agent are autonomous. The AI agent gives hints, and the human needs to guess the word. The speed of the task is low due to ambiguity in possible guesses, which leads to uncertainty and slower decision-making. The complexity is high as each hint affects subsequent choices, creating a hierarchical and complex task. The importance of the task is high because each step is irreversible. Once a hint or guess is made, it cannot be undone.

Carroll et al. (2020) explore a cooperative digital cooking environment where both the human and AI agent work to prepare and serve a dish autonomously. The speed of this task is high due to the clear and unambiguous outcomes of each step. The task is relatively simple with low complexity because the steps do not influence each other, making it non-hierarchical. However, the importance remains high because actions as cooking something and serving a dish are irreversible in this cooking game.

Dell'Acqua et al. (2023a) focus on cooperative ingredient selection. Both the human and the AI agent are autonomous and have the same shared objective of completing as many recipes as possible within one minute by picking the correct ingredients. The speed of this task is high as there is no ambiguity. The outcomes of each step they take, picking ingredients, are clear. Furthermore the task is not complex. The steps they take in this task do not influence other steps. They can always continue to pick the ingredients they want. Lastly, the importance of this task is also low. If an incorrect ingredient is picked, it can be put back.

Fügener et al. (2021) examine cooperative image classification where the human and AI agent can delegate classification tasks to each other. The task speed is low as ambiguity arises due to uncertainty in outcomes of classifying images. This requires both humans and AI agents to carefully consider their decisions. Moreover, the task is hierarchical. For example when the human or the AI agent decides to classify an image themselves, the option to delegate the classification work to the other is not possible anymore. This makes the execution of the task more complex. In addition, the task is very important since the execution of the task is irreversible. Once a classification decision or decision to delegate is made it is not possible to reverse that decision.

Gnewuch et al. (2023) experiment in a digital chat environment for service assistance where the human and AI agent cooperate. The task is moderately ambiguous since the outcomes of steps can be uncertain, slowing down execution. Furthermore, the task is non-hierarchical, as the outcome of one step does not affect the choices available in subsequent steps. Therefore, the task is not complex. Additionally, the task is of low importance as the steps are reversible. For instance, the human can send a new message to the AI agent on a different topic.

Han et al. (2022) focus on a cooperative digital chat environment addressing product order issues. There is no form of ambiguity in this task as the outcomes of various steps are clear. The human knows, based on a cover story, the correct response to the message of the AI agent, and the AI agent messages are predetermined. This means that the task will be executed fast. The complexity is low since the task is non-hierarchical and predetermined steps do not influence each other. However, the task is important because the steps taken are irreversible.

Li et al. (2022) focus on a cooperative digital game where humans and the AI agent need to destroy an opponent's base. The task speed varies as some steps have uncertain outcomes, which can slow execution. The complexity is high because the outcome of one step often affects subsequent options, making the task hierarchical. The importance is also high because the steps are irreversible.

J. Liu et al. (2024) study a collaborative digital cooking environment where the human gives instructions to the AI agent. The task speed is high as outcomes are clear and unambiguous. The complexity is low because the steps do not influence each other, allowing the task to remain non-hierarchical. Both can always start from the beginning of the process and prepare and serve the dish. The task is important because steps as serving a dish cannot be undone once completed, highlighting the importance of each step of the task.

Finally, Lucas et al. (2014) look at a cooperative virtual health interview setting where both the human and the AI agent aim to complete a medical interview successfully. The speed is high due to the certainty of outcomes. The question of the AI agent will always lead to an answer of the human, and an answer always prompts a new question. Additionally, this task is not hierarchical. The response of the human for example does not limit the AI agent in its subsequent questions and the human can always provide answers without being influenced by other parts of the task. The task is important as steps of the task are irreversible.

2.3.2 Findings

Past research has explored a wide range of application areas of Human-AI interactions. These areas include game environments such as cooking or word prediction environments, as well as chat, spoken interaction interviews, and image classification environments. What can be observed is that all these environments in the previous studies are digital spaces where the interactions are increasingly prevalent and observable. These digital spaces are controlled environments, which means that they have specific boundaries within which humans and AI agents operate. The boundaries define the scope of possible actions and interactions, ensuring a consistent and predictable context for both humans and AI agents.

Past research primarily focuses on two themes within these digital controlled environments. The first theme explores the importance of designing AI agents specifically for human interactions to improve the performance of the interactions. The second theme examines how human perceptions and behaviours differ in interactions with AI agents compared to interactions with other humans.

Several of the previous studies show that not all AI agents are suitable for interactions with humans (Carroll et al., 2020; J. Liu et al., 2024). This leads to the emergence of the first theme that explores the importance of designing AI agents specifically for human interactions. Carroll et al. (2020) show that AI agents that are designed for interactions with humans perform much better compared to those not specifically designed for these interactions. The difference is that AI agents specially designed for human interactions no longer assume that they perform interactions with the perfect partner and are more flexible in their interactions. In addition, J. Liu et al. (2024) show that a combination of different models behind the AI agent is more effective than using a single model. Combining a slower but highly accurate Large Language Model (LLM), a lightweight LLM and a model executing predetermined actions outperforms using individual models, resulting in better interaction outcomes. The AI agent has more different skills, including the ability to reason well with the slow but accurate model, but also being able to perform fast interactions with the lightweight LLM. This highlights that not all AI agents are suitable for interaction with humans, showing the importance of selecting an AI agent specifically designed for these interactions.

The second theme emerges from studies that examine positive and negative sides of human perceptions and behaviours towards AI agents in interactions. A positive side of Human-AI interactions is that humans feel more at ease than when they interact with other humans. According to Lucas et al. (2014), humans often show reluctance in sharing personal information, particularly in medical contexts, where they strive to present a more favourable impression. However, when interacting with an AI agent instead of a human, this reluctance diminishes along with the inclination to create a positive impression. To achieve this it is also important that the AI agent resembles a human. Vanneste and Puranam (2024) argue that humans have more trust in AI agents in that situation. When humans perceive an AI agent as more trustworthy, cooperative interactions result in enhanced performance (Li et al., 2022).

However, there are studies that contradict these findings and show that interactions between humans and AI agents do not only have positive aspects. Vanneste and Puranam (2024) argue that because of humans' aversion to betrayal, the fear of the psychological impact of breaching trust by the AI agent intensifies when it becomes more human-like. Moreover, several studies show that humans perceive AI agents differently than humans (Ashktorab et al., 2020; Dell'Acqua et al., 2023a; Fügener et al., 2021; Gnewuch et al., 2023; Han et al., 2022). Even if the AI agent provided interactions identical to those of human participants during interactions with other humans, they still perceived the AI agent as having less abilities while this is not the case (Ashktorab et al., 2020). In addition, Dell'Acqua et al. (2023a) show that replacing a human with an AI agent can lead to a decline in team performance and even failures in the interaction. This often results in diminished trust, even when the AI agent performs the task more effectively than the replaced human. When humans and AI agents have to interact, humans prefer interactions where they both work together and where their actions affect the responses of the other and do not prefer interactions where they work separately (Li et al., 2022).

The same holds for mistakes made by humans and AI agents. In the case of delegation, AI agents positively influence performance when they delegate. However, when an AI agent makes a mistake, even if it is smaller than a mistake made by a human, humans react more intensely to that mistake (Fügener et al., 2021). This also applies to emotions expressed by humans and AI agents as humans are more satisfied with a service when a human agent expresses positive emotions in their interactions than when such emotions are expressed in interactions with an AI agent (Han et al., 2022). In addition, communication from humans to AI agents is completely different than to other humans. Humans are more likely to communicate in a human way when they know human involvement is revealed in interactions with service agents. This means they use longer and more detailed messages instead of short, simple keywords, because they want to leave good impressions to other humans (Gnewuch et al., 2023).

2.4 Rationale for an Experimental Approach to Study Human-AI Interactions

This demonstrates that numerous studies have already been conducted on Human-AI interactions, with varying outcomes across two main themes. However, there are many other themes within these interactions that remain to be explored. To explore more about Human-AI interactions, an experimental approach is well suited for two reasons. Firstly, these interactions occur in digital spaces, where they are increasingly prevalent and observable. Digital environments are well-suited for experiments because they allow for precise control over variables and systematic observation of participant behaviour. By conducting experiments in these digital spaces, researchers can meticulously observe how humans respond to AI in interactions and vice versa. Secondly, the interactions take place in controlled environments, where humans and AI agents interact within defined boundaries. These boundaries outline possible actions and interactions, creating a consistent and predictable environment for participants of the experiment. These environments are extremely suitable for experiments because they provide researchers with the ability to conduct experiments under specific conditions, allowing for precise observation of the effects of various factors. Maintaining consistent boundaries allows researchers to establish a stable and predictable setting for Human-AI interactions, which enhances the reliability of the results.

3 Experimental Design

3.1 Objective of Experiments

The primary objective of experimentation is to elucidate causal relationships (Levine et al., 2023). It entails investigating how particular factors influence a given response or outcome. This is achieved by identifying the independent variable(s) and observing their impact on the dependent variable(s) (Berger et al., 2018).

3.2 Steps in Designing an Experiment

Designing an experiment typically involves four main steps. The first step entails determining the type of experiment to conduct. This type includes a laboratory or a field experiment, or alternatively, a non-experimental method (Levine et al., 2023). In laboratory and field experiments, manipulation of the independent variable occurs to examine its effects, whereas such manipulation is absent in non-experimental methods (Jhangiani et al., 2019). According to List (2007), in laboratory experiments, control over the laboratory environment is imperative to accurately measure the effects of treatments. Through establishment of a control group through randomisation, it provides a robust method to establish causality within the experimental design. Field experiments similarly employ randomisation but are conducted within natural settings, seamlessly integrating with participants' regular activities (List, 2007). Second, it is necessary to define a precise experimental situation. This involves outlining the independent and dependent variables, as well as defining the procedures. These procedures include establishing the sequence where participants are exposed to different conditions or treatments while also incorporating a cover story (Levine et al., 2023). The third step involves pre- and pilot testing, along with pre registration. Running a pilot test of the complete experiment is crucial to detect early potential problems, enhance methods and tools, and assess the practicality of the experiment. Pre registration involves outlining research methods and tools before starting with the experiment (Levine et al., 2023). The last step is collecting the data, doing a follow-up and analysing the data (Levine et al., 2023). This allows for a thorough evaluation of the research hypothesis and clear conclusions about the relationships between the variables under investigation, depending on the chosen research method.

4 Human-AI Interaction Experimental Guidelines

These guidelines form a robust basis for conducting experiments on Human-AI interactions. Building on these guidelines, we suggest additional considerations for the design of Human-AI interactions. These considerations are based on the three dimensions of interaction tasks and the three different interaction types. We also consider what requirements the platform for the interaction experiment must meet.

4.1 Dimensions of Interaction Tasks

We start by considering the different dimensions of the interaction task and propose different implementations of experiments. Table 2 presents the three different dimensions of tasks: speed, complexity and importance, accompanied by their experimental approach and requirements for successful interaction. The experimental approach contains experimental examples and expectations from humans. Each dimension is examined in terms of its high and low aspects, allowing a successful Human-AI interaction experiment to be set up for any type of task.

Table 2

Different Task Dimensions and Their Experimental Approach and Requirements for Successful Interaction

Task dimension	Experimental approach	Requirements for successful interaction
Speed (Low)	Tasks with a slower pace due to ambiguity of the task, because outcomes of several steps of the task are unclear. Experimental example *Word guessing task* [2] Expectations from humans Humans need to carefully consider different options beforehand to make well-informed decisions about the steps they will take.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The model to use for the AI agent should be able to careful consider each step in the task and be able to determine the best step to take.

[2] The human and the AI agent interact to guess as many words correctly as possible, where the human provides hints and the AI agent needs to guess the word

Task dimension	Experimental approach	Requirements for successful interaction
Speed (High)	Task with faster pace because there is no ambiguity. The outcomes of several steps of the tasks are clear. Experimental example *Cooking task* [3] Expectations from humans There are no direct expectations from humans as the clear outcomes allow for quick and efficient execution.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The AI agent should be based on a model capable of handling fast-paced task execution.
Complexity (Low)	Tasks with little hierarchy, the steps in performing the task are not dependent on each other. Experimental example *Cooking task* [3] Expectations from humans There are no direct expectations from humans since the steps are independent and can be performed independently.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The model to use for the AI agent does not need to handle high complexity and can focus on executing straightforward, independent tasks efficiently.

[3] The human and the AI agent have to prepare and serve dishes in a cooking environment

Task dimension	Experimental approach	Requirements for successful interaction
Complexity (High)	Tasks with a lot of hierarchy, one step in executing a task has a significant impact on the available options for another step of executing the task Experimental example *Word guessing task* [4] Expectations from humans Humans need to carefully consider the consequences of each step as it influences the choices and possibilities of following steps.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The AI agent should be based on a model that can handle complex tasks.
Importance (Low)	Tasks with low stakes, where errors have minimal consequences due to reversibility. Experimental example *Drawing task* [5] Expectations from humans There are no direct expectations from humans since the steps are reversible and they can make mistakes.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The AI agent should use a model that is simple, focusing on efficiency and error recovery.

[4] The human and the AI agent interact to guess as many words correctly as possible, where the human provides hints and the AI agent needs to guess the word. The human is allowed to provide one word as a hint at a time, after which the AI agent can guess one word. If the guess is incorrect, the human can provide another hint.

[5] The human and the AI agent play a drawing game. One draws a word, and the other needs to guesses the correct word while the other is drawing.

Task dimension	Experimental approach	Requirements for successful interaction
Importance (High)	Tasks with high stakes, where errors can have significant consequences due to the irreversibly. <u>Experimental example</u> *Word guessing task* [6] <u>Expectations from humans</u> Humans need to carefully consider whether they want to take a step in the task since it cannot be reversed.	Determine the actions of the human and the AI agent based on the chosen interaction type (refer to table 3). The model to use for the AI agent should ensure high accuracy and reliability, to minimize the risk of mistakes in crucial steps which are irreversible.

[6] The human and the AI agent interact to guess as many words correctly as possible, where the human provides hints and the AI agent needs to guess the word. The human is allowed to provide one word as a hint at a time after which the AI agent tries to guess the word.

4.1.1 Model for the AI Agent

It is important to acknowledge the significant differences in the effectiveness of AI models (X. Liu et al., 2023). This underlines the need to carefully select an appropriate model based on the task the Human-AI interaction intends to accomplish within the experiment. Selecting an LLM in combination with other models, as the foundation for the AI agent, could offer valuable advantages according to X. Liu et al. (2023). Focusing primarily on practical tasks, LLMs are increasingly demonstrating greater intelligence and autonomy (X. Liu et al., 2023) and possess robust logic (J. Liu et al., 2024). However, what type of model to prefer is different for each task in the Human-AI interaction experiment.

4.2 Types of Interactions

In addition, we consider the three different types of Human-AI interactions (Schmidt & Loidolt, 2023) and propose different implementations of experiments. The various Human-AI interaction types, i.e., coordination, collaboration and cooperation, all entail distinct experimental approaches. This includes unique experiment setups, research questions, expected human behaviour and requirements for achieving successful interactions. Table 3 contains the different types of interactions, accompanied by their experimental approach and requirements for successful interaction. The experimental approach consists of three parts, the experimental examples, the questions that could be answered and the expectations from humans.

Table 3

Different Interaction Types and Their Experimental Approach and Requirements for Successful Interaction

Interaction Type	Experimental approach	Requirements for successful interaction
Coordination	Experimental example *In a virtual maze, both humans and the AI agent have to find their own exit. At the same time they can place obstacles to hinder each other's progress. While navigating, they can predict each other's movements and the obstacles in the environment and adapt strategies to optimise their progress.* Questions that could be answered • How do humans and AI agents adapt their behaviour when pursuing independent goals in the same environment? • How accurately can humans anticipate the actions of the AI agent? • How effectively can the AI agent interpret human actions? Expectations from humans Human participants are expected to understand and adapt to both their own objectives and those of the AI agent in the shared environment. They should effectively coordinate with the AI agent, adjusting their strategies when needed.	Establish an operational context and design framework for the AI agent: • Determine the type of information the AI agent can comprehend. • Outline how the AI agent should act in various situations. Actions of the human and the AI agent: • Enable humans to recognise and interpret the AI agent's actions. • Ensure the AI agent can recognise and interpret human actions. (Schmidt & Loidolt, 2023). Take in consideration how the dimensions of the task affect the interaction between the human and the AI agent.

Interaction Type	Experimental approach	Requirements for successful interaction
Collaboration	Experimental example *In a virtual cooking environment, humans and the AI agent must collaborate to prepare a meal. Each of them has access to the ingredients and utensils. They need to interact together to create the meal. The human needs to instruct the AI precisely on which tasks it needs to perform.* Questions that could be answered • How do participants communicate and collaborate their actions toward a shared objective? • To what extent does the AI agent rely on human instructions? • How does the interaction between humans and AI evolve as they work together? Expectations from humans Human have to identify and recognise the AI agent and give clear instruction to AI agent on what to do. This is crucial for seamless collaboration.	Establish the operational context and design framework. [7] Actions of the human and the AI agent: • Identify and recognise each other. • Recognise and interpret each other's actions aimed at achieving the shared objective. • It is imperative for the AI agent to consistently follow human instructions and operate as a tool to accomplish human objectives. (Schmidt & Loidolt, 2023). Take in consideration how the dimensions of the task affect the interaction between the human and the AI agent.

[7]Similar to the coordination interaction type.

<table>
<tr><th>Interaction Type</th><th>Experimental approach</th><th>Requirements for successful interaction</th></tr>
<tr><td>Cooperation</td><td><u>Experimental example</u>
In a virtual planning simulation, humans and the AI agent must cooperate to design an efficient public transportation network. They are tasked with planning and optimising subway routes. The AI agent can make autonomous decisions, while humans can also introduce their own strategies to solve the problem.
<u>Questions that could be answered</u>
• How do participants coordinate their actions toward a shared objective when the AI agent has autonomy?
• To what extent can the AI agent generate useful insights independently?
• How does the distribution of decision-making authority between humans and AI impact collaboration?
<u>Expectations from humans</u>
Human participants are expected to identify and acknowledge the AI agent. This entails comprehending their actions and interacting actively to achieve the common objective.</td><td>Establish the operational context and design framework. [8]
Actions of the human and the AI agent:
• Capable of identifying and acknowledging each other.
• Recognise and interpret the actions of each other.
• The AI agent can generate its own insights and information.
– It should not solely rely on instructions provided by humans.
(Schmidt & Loidolt, 2023).
Take in consideration how the dimensions of the task affect the interaction between the human and the AI agent.</td></tr>
</table>

[8]Similar to the coordination interaction type.

4.3 Platform Requirements

When conducting experiments related to those Human-AI interactions, having a suitable platform is crucial for facilitating and observing those interactions in a controlled and efficient manner. Such a platform needs to serve as the operational environment where researchers can systematically manipulate variables, observe interactions, and draw conclusions. The platform should provide a controlled environment where it is possible to manipulate variables and observe interactions under controlled conditions. This allows for precise experimentation and reliable results. Researchers often have specific requirements for their experiments. Therefore, the platform should be customisable, allowing researchers to tailor the environment and parameters according to their experimental design. The platform should also be scalable to accommodate various levels of interaction, from small-scale interactions between individuals to large-scale interactions involving multiple groups. This scalability enables researchers to conduct experiments of different sizes and complexities. In addition, the platform should ensure the privacy and security of participants' data safety.

5 Discussion

The primary aim of this thesis is to establish guidelines for designing experiments to study Human-AI interactions. As AI technology becomes increasingly integrated into different aspects of society, understanding how to conduct research into these interactions is crucial. This thesis shows that past research conceptualised Human-AI interactions based on different forms of relation between humans and AI agents. The different interaction types are coordination, collaboration, and cooperation. We underscore the importance of considering this perspective of the parties involved, but also the task-oriented perspective where we borrow from the strategic decision-making literature to conceptualise the task along three different dimensions: the speed, complexity, and importance of tasks, to be able to provide a better understanding of the interactions. Current studies already indicate that setting up an experiment involves four key steps. First, determining the type of experiment to conduct. In addition, defining a precise experimental situation. Third, conducting pre- and pilot testing, along with pre registration. Finally, collecting the data, performing a follow-up, and analysing the data. In addition we examine the two perspectives to offer a comprehensive guideline specifically for designing Human-

AI interactions experiments. We examine the experimental approach and requirements for each interaction type and task dimension as they all entail distinct experimental approaches. We show that the type of model to use for the AI agent is different per task. This ensures that each type of interaction and task can be effectively implemented within an experiment. This leads to more successful interactions and thus achieves better experimental results. Finally, it is also essential that the platform on which the experiment is conducted meets various requirements.

5.1 Contributions

We contribute to the literature on Human-AI interaction (Choudhary et al., 2023; Schmidt & Loidolt, 2023). We do this by addressing a critical gap in the existing literature, the lack of clear guidelines for designing Human-AI interaction experiments. By creating these guidelines, this thesis offers a structured framework that categorises Human-AI interactions by the perspective of the parties involved and task-oriented perspective. Using these two perspectives we can enhance our understanding for setting up experiments specifically designed for these Human-AI interactions. Before this thesis, there were no guidelines on how to design these experiments. We now have a clearer understanding of the specific experimental approaches and requirements for successful interaction needed for the different task dimensions and the different types of interactions. The importance of this contribution is very high. As human-AI interactions become more common, there is a growing need for well-designed experiments to understand and improve these interactions. The guidelines in this thesis will help researchers create better experiments, which will lead to more accurate and useful results. Future research should focus on validating the proposed guidelines for setting up Human-AI interaction experiments in various contexts.

We also contribute to the literature on strategic decision-making (Zohrehvand, 2020). This thesis contributes to the literature on strategic decision-making by applying its concepts to study Human-AI interactions. Specifically, we use the three main characteristics of strategic decision-making to conceptualise the task the interactions intend to accomplish to provide a more comprehensive understanding of these interactions. This application helps connect strategic decision making with Human-AI interaction research, offering a new way to analyse these interactions. We now know that the integration of strategic decision-making concepts provides a more detailed understanding of how these interactions are structured based on their task. This was not highlighted in past research, which often focused only on the relation between the interacting parties. This understanding is important because it offers a more complete framework understanding and designing Human-AI interactions for experiments. Future research should explore how these dimensions impact different types of interactions differently, enabling a clearer differentiation between interaction types. This differentiation would help participants better understand what to expect and facilitate more seamless interactions, ultimately leading to better results.

5.2 Practical Implications

The guidelines presented in this thesis for designing experiments to study Human-AI interactions have significant practical implications. These implications are relevant for both researchers and participants in the experiments. For researchers, the detailed guidelines mean improved experimental designs. By following the specific guidelines for different types of interactions, coordination, collaboration, and cooperation and using the task-oriented approach that considers the speed, complexity, and importance of tasks, researchers can set up more robust and accurate experiments. Researchers can also choose the most appropriate AI model that fits the specific task and know what the requirements are for a platform for their Human-AI interaction experiment. This leads to more reliable results and deeper insights into how humans and AI agents interact together. Participants of the experiment also benefit from these improved experimental designs. Because the experiment has a better design and fits well with the interaction type and the task the interaction intends to accomplish, the interaction will proceed in a more natural and intuitive way. This reduces possible frustrations and increases participant involvement.

5.3 Limitations

Despite the comprehensive guidelines provided in this thesis, several limitations must be acknowledged. Firstly, our research was mainly limited by the available literature, which might not cover all possible forms of Human-AI interactions. While we focused on coordination, collaboration, and cooperation, there might be other interaction types that were not considered, which could limit the generalisability of our findings. Future research can address this by exploring and identifying more types of Human-AI interactions, creating a more inclusive framework. Additionally, the task-oriented perspective relies on the dimensions speed, complexity, and importance. This categorisation might not fully capture the details of all tasks in different experimental settings. To overcome this limitation, future research should aim to refine and expand the task-oriented framework, incorporating more dimensions to capture a broader range of task characteristics. It is also a limitation that it has not been distinguished how the three dimensions of the task influence each interaction type differently. Future research can further investigate how these dimensions affect the various interaction types in different ways to overcome this limitation. Lastly, our guidelines were developed based on theoretical and experimental studies, but their application in real-world settings remains to be thoroughly validated. Future research should aim to test these guidelines in various settings to ensure their robustness and applicability across different domains of Human-AI interactions.

5.4 Conclusion

In conclusion, this thesis provides guidelines for designing Human-AI interaction experiments by integrating both relational and task-oriented perspectives. By categorising interactions into coordination, collaboration, and cooperation, and examining tasks through the dimensions of speed, complexity, and importance, we offer a comprehensive approach for setting up reliable experiments. These perspectives not only enhance the understanding of Human-AI interactions but also bridges a critical gap in the existing literature. Future research should continue to validate and expand these guidelines across diverse contexts, ensuring they remain relevant and effective in studying the increasingly prevalent interactions between humans and AI agents.

References

Angus, S., & Christine, L., A. (2010). *Oxford Dictionary of English, 3rd Edition.* Oxford University Press.

Ashktorab, Z., Liao, Q. V., Dugan, C., Johnson, J., Pan, Q., Zhang, W., ... Campbell, M. (2020, oct). Human-AI Collaboration in a Cooperative Game Setting: Measuring Social Perception and Outcomes. *Proc. ACM Hum.-Comput. Interact.*, *4*(CSCW2). Retrieved from https://doi.org/10.1145/3415167 doi: 10.1145/3415167

Berger, P., Maurer, R., & Celli, G. (2018). *Experimental Design with Applications in Management, Engineering, and the Science - 2nd Edition.* doi: 10.1007/978-3-319-64583-4

Carroll, M., Shah, R., Ho, M. K., Griffiths, T. L., Seshia, S. A., Abbeel, P., & Dragan, A. (2020). *On the Utility of Learning about Humans for Human-AI Coordination.* Retrieved from https://doi.org/10.48550/arXiv.1910.05789

Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2023). Human-AI Ensembles: When Can They Work? *Journal of Management*, *0*(0). Retrieved from https://doi.org/10.1177/01492063231194968 doi: 10.1177/01492063231194968

Dell'Acqua, F., Kogut, B., & Perkowski, P. (2023a). Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination. *The Review of Economics and Statistics*, 1-47. Retrieved from https://doi.org/10.1162/rest_a_01328 doi: 10.1162/rest_a_01328

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... Lakhani, K. R. (2023b). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. Retrieved from http://dx.doi.org/10.2139/ssrn.4573321 doi: 10.2139/ssrn.4573321

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.adn5290 doi: 10.1126/sciadv.adn5290

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678-696. Retrieved from https://doi.org/10.1287/isre.2021.1079 doi: 10.1287/isre.2021.1079

Gnewuch, U., Morana, S., Hinz, O., Kellner, R., & Maedche, A. (2023). More Than a Bot? The Impact of Disclosing Human Involvement on Customer Interactions with Hybrid Service Agents. *Information Systems Research*, *0*(0). Retrieved from https://doi.org/10.1287/isre.2022.0152 doi: 10.1287/isre.2022.0152

Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior Research Methods*, *52*(6), 2283–2286. Retrieved from http://dx.doi.org/10.3758/s13428-020-01395-3 doi: 10.3758/s13428-020-01395-3

Han, E., Yin, D., & Zhang, H. (2022). Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service? *Information Systems Research*, *34*(3), 1296-1311. Retrieved from https://doi.org/10.1287/isre.2022.1179 doi: 10.1287/isre.2022.1179

Hornbæk, K., & Oulasvirta, A. (2017). What Is Interaction? In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 5040–5052). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025765 doi: 10.1145/3025453.3025765

Jhangiani, R., Chiang, I.-C. A., Cuttler, C., & Leighton, D. C. (2019). *Research methods in psychology* (4th ed.). Kwantlen Polytechnic University.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015, 12). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, *34* (Suppl), 1220–1228. Retrieved from http://dx.doi.org/10.1037/hea0000305 doi: 10.1037/hea0000305

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments.* Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/9781108653985 doi: 10.1017/9781108653985

Levine, S. S., Schilke, O., Kacperczyk, O., & Zucker, L. G. (2023). Primer for Experimental Methods in Organization Theory. *Organization Science*, *34* (6), 1997-2025. Retrieved from https://doi.org/10.1287/orsc.2023.18093 doi: 10.1287/orsc.2023.18093

Li, J., Huang, J., Liu, J., & Zheng, T. (2022). Human-AI cooperation: Modes and their effects on attitudes. *Telematics and Informatics*, *73*, 101862. Retrieved from https://www.sciencedirect.com/science/article/pii/S0736585322000958 doi: https://doi.org/10.1016/j.tele.2022.101862

List, J. A. (2007). Field Experiments: A Bridge between Lab and Naturally Occurring Data. *The B.E. Journal of Economic Analysis Policy*, *6* (2), 0000102202153806371747. Retrieved 2024-07-12, from https://doi.org/10.2202/1538-0637.1747 doi: doi:10.2202/1538-0637.1747

Liu, J., Yu, C., Gao, J., Xie, Y., Liao, Q., Wu, Y., & Wang, Y. (2024). *LLM-Powered Hierarchical Language Agent for Real-time Human-AI Coordination.* Retrieved from https://arxiv.org/abs/2312.15224

Liu, X., Yu, H., Zhang, H., Xu, Y., Lei, X., Lai, H., . . . Tang, J. (2023). *AgentBench: Evaluating LLMs as Agents.* Retrieved from https://arxiv.org/abs/2308.03688

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer:

Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94-100. Retrieved from https://www.sciencedirect.com/science/article/pii/S0747563214002647 doi: https://doi.org/10.1016/j.chb.2014.04.043

McKinsey. (2022). *The state of AI in 2022—and a half decade in review.* Retrieved from https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai%20in%202022%20and%20a%20half%20decade%20in%20review/the-state-of-ai-in-2022-and-a-half-decade-in-review.pdf?external_link=true

Poole, L. D., & Mackworth, K. A. (2010). *Artificial Intelligence: Foundations of Computational Agents*. USA: Cambridge University Press.

Schmidt, P., & Loidolt, S. (2023). Interacting with Machines: Can an Artificially Intelligent Agent Be a Partner? *Philosophy and Technology*, *36*(3). Retrieved from http://dx.doi.org/10.1007/s13347-023-00656-1 doi: 10.1007/s13347-023-00656-1

Tanco, M., Costa, N., & Viles, E. (2009). Experimental design selection: guidelines for practitioners. *International Journal of Productivity and Quality Management*, *4*(3), 283. Retrieved from http://dx.doi.org/10.1504/IJPQM.2009.023698 doi: 10.1504/ijpqm.2009.023698

Vanneste, B. S., & Puranam, P. (2024). Artificial Intelligence, Trust, and Perceptions of Agency. *Academy of Management Review*. Retrieved from http://dx.doi.org/10.5465/amr.2022.0041 doi: 10.5465/amr.2022.0041

Wang, W., Gao, G. G., & Agarwal, R. (2023). Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience. *Management Science*. Retrieved from https://doi.org/10.1287/mnsc.2021.00588 doi: 10.1287/mnsc.2021.00588

Webster, M., & Sell, J. (2014). Chapter 1 - Why Do Experiments? In M. Webster & J. Sell

(Eds.), *Laboratory Experiments in the Social Sciences (Second Edition)* (Second Edition ed., p. 5-21). San Diego: Academic Press. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780124046818000017 doi: https://doi.org/10.1016/B978-0-12-404681-8.00001-7

Zohrehvand, A. (2020). M&As and CEOs: Machine Learning Aided Analyses of Social Media. Retrieved from https://discovery.ucl.ac.uk/id/eprint/10117933/