Understanding Users' Contextual Factors and Personal Values for Watching YouTube Videos:

A Crowdsourcing Approach with Personal Reflection Integration

Master Thesis

Yizhen Zhang

Delft University of Technology

TU Delft

Understanding Users' Contextual Factors and Personal Values for Watching YouTube Videos:

A Crowdsourcing Approach with Personal Reflection Integration

by

Yizhen Zhang

Student Name	Student Number
Yizhen Zhang	5801133

Project Duration:	November, 2023 - September, 2024	
Faculty:	Faculty of Electrical Engineering, Mathematics and Computer Science, Delft	
Thesis Committee:	Dr. ir. Ujwal Gadiraju,	TU Delft, Chair
	Dr. ir. Jie Yang,	TU Delft, Daily Supervisor
	Dr. ir. Huijuan Wang,	TU Delft, Core Member
	Msc. Di Yan,	TU Delft, External Member

TU Delft

Acknowledgement

As I write these words, I realize that my journey as a graduate student is nearing its end. A sense of relief fills my heart, yet mingling with faint sadness. These two short years, which began in late summer and now conclude in early fall, have flown by like a fleeting moment. Looking back, there were times of pressure, joy, sadness, and loneliness. Yet all of these emotions have quietly dissolved into the ordinary rhythm of daily life. Now, in the final moments before graduation, I must bid farewell in a rush, leaving behind endless memories. I hope, in the future, when I embark on new adventures, I will reminisce in the quiet of the night about the streets lined with cherry blossoms during springtime, the gusts of wind that greet me as I pass by the towering EWI building, the sunlight warming the library lawn, the ducks swimming in the pond, and every corner of TU Delft filled with either laughter or tranquility that once brought me comfort.

As I reflect on these cherished memories, I realize that none of this journey would have been possible without the support and guidance of many people. First and foremost, I would like to thank my thesis committee. Dr. Jie Yang provided an invaluable and promising topic and gave me the opportunity to work with the support of outstanding groups. During the whole process, he offered excellent insights into the direction of my research and thesis. Msc. Di Yan helped me a lot in the parts of design and analysis. The weekly meeting with her always provided me with new inspirations and clear directions. She also encouraged me sincerely and took care of my emotions when I was depressed and overwhelmed. Zihan Wang was a great collaborator. Since we were doing similar research, we shared our ideas with each other and offered help when the other was in trouble. He assisted me with several technical issues during my research.

Additionally, I would appreciate my friends both in China and in the Netherlands. They were constant companions and strength throughout my journey and were always willing to offer me a listening ear. Whenever I was in moments of joy or sorrow, I could rely on them to share my experiences and lift my spirits. Their company reminded me that I was never alone.

Most importantly, a big thank you to my family. Starting life in a country with a completely different culture was not easy, but my parents were always giving me unconditional love and encouragement. During the time I was completing my Master’s Thesis Project, I faced many other pressures as well. During the countless nights when it was difficult to sleep or when I was confused and at a loss for words, my parents were always my strong backing, accompanying me and unwavering in their support of all the decisions I made.

There are countless people who have been present on my journey. Now I take this opportunity to thank those who have accompanied me along the way and who have provided me with encouragement, guidance, and love during these years.

Yizhen Zhang
Delft, September 2024

Abstract

User feedback plays a significant role in helping recommendation systems to make personalized and accurate predictions. Despite the fact that many methods of collecting user feedback have been proposed, little research exists that addresses both the breadth and depth of data collected. In this study, we incorporate personal reflection into traditional crowdsourcing tasks and investigate how it facilitates people's reflection on their usage of YouTube. We present a novel crowdsourcing approach with personal reflection integration based on several design principles, which allows participants to reflect on contextual factors and personal values of using YouTube through guided context recall and evaluations, therefore gathering deeper insights into people's preferences and behaviors on a large scale. We conducted a user study involving 20 participants and explored the insights generated from their textual answers and the role of design principles in the reflection process. This approach successfully enables multiple participants to conduct the study simultaneously, thereby reflecting on their watching behaviors and preferences. Based on the quantitative and qualitative analysis of the findings, we sum up the implications of this approach to provide guidance for the YouTube recommendation system and point to the directions for the design of similar studies in the future.

Contents

1 Introduction

In the era of digital information and user-centric online platforms, recommendation systems serve as an indispensable tool, guiding users to navigate a wide variety of content and recommending items, services, and products based on their preferences and behaviors. The user preferences are derived from user feedback [56]. By analyzing user feedback such as direct comments and previous interactions, these systems are able to identify associations among large amounts of information and provide personalized content, ultimately enhancing user engagement and experience [61]. Consequently, there is a research need for recommendation systems to gather and explore rich user feedback, encompassing both the depth and breadth of the collected data, to optimize the recommendation algorithms, therefore improving the quality of recommendations to make more accurate predictions. However, most existing recommendation systems fall short of fulfilling this demand. In most cases, the recommendation system appears as a “black box”, which means it prevents users from comprehending recommended results and providing feedback [27]. Most importantly, there are currently no studies that address both the depth and breadth of data collected, which means the collected feedback used for recommendations is either shallow in quality or insufficient in quantity. To be specific, the shallowness of feedback lies in the inability of the systems to have a comprehensive understanding of the reasons behind users' preferences for an item. Most current methodologies of collecting user feedback lack the capability to delve into the exact reasons that drive users to either like or dislike a particular item. Instead, they narrowly concentrate on limited types of feedback, such as numerical scales or simplistic binary indicators like ”thumb-up” and ”thumb-down”. These approaches restrict users from expressing their opinions, resulting in superficial understandings that miss the underlying motivations, contextual factors, and individualized considerations that contribute to user preferences.

Recognizing the limitations of the existing user feedback mechanisms, we need to gather user feedback that goes beyond surface-level reactions. Most research on recommendation systems typically focuses on the accuracy of their prediction algorithms [43]. However, accuracy is only one aspect of the overall user experience. There are many other factors that significantly influence user experience. In our research, we merely focus on the users' contextual characteristics and personal values [43] for recommendation systems. One possible solution is to employ interviews. However, interviews are usually conducted on a one-on-one basis, which requires a large amount of the investigator's time [2]. Interviews are therefore typically performed on a small scale, resulting in the inability to obtain a sufficient amount of user feedback.

In order to collect sufficient data while going deep enough, the concept of crowdsourcing emerges as a powerful paradigm to address the demand. It leverages collective intelligence, particularly the ability of different individuals, to collectively reach various goals [24], ranging from simple data annotation to problem-solving and innovation. Over the past decade or so, it has shown to be a promising approach to addressing some of the growing challenges associated with data collection [80]. By leveraging the collective intelligence of a diverse crowd, we are able to deploy surveys or interactive tasks to gather open-response feedback [22]. For example, a crowdsourced online feedback platform is an emerging mechanism for gathering large amounts of feedback quickly [22] on specific experiences or preferences.

These crowdsourcing approaches can probably be used to refine the effectiveness of recommendations [62] and provide a more personalized and enriched user experience. In consequence, we investigated the possibility of utilizing the concept of crowdsourcing to fulfill the need of recommendation systems to gather diverse insights and perspectives about user experiences.

Nevertheless, the current landscape of crowdsourcing methods and tools encounters a significant challenge. The tasks presently available on crowdsourcing platforms often lack the depth and engagement required for insightful data collection. Their simplistic designs limit participants' ability to provide valuable insights, resulting in shallow insights and limited participant engagement. To be specific, the majority of existing crowdsourced data management works are focused on micro-tasks [50], which require only a brief period for completion. They focus on basic data representation and simple surveys, failing to fully capture the complexity of user preferences and behaviors. Moreover, current crowdsourcing tasks are characterized as repetitive with little diversity. This is likely to cause monotony and participant fatigue, ultimately resulting in decreased worker engagement and risking the output quality [76]. These issues highlight the necessity for more sophisticated and appealing data and tasks that can have positive influences on the participant's engagement and willingness to conduct the study properly [8]. Besides these limitations, the research on the overlapping areas of recommendation systems and crowdsourcing is presently underexplored as well. Existing literature predominantly focuses on the work of utilizing ad-hoc feedback [25] to improve the performance of recommendation systems, or micro-tasks to gather simplistic and brief feedback, yet lacks a systematic investigation of how crowdsourced data can enrich the effectiveness of recommendation algorithms. Based on the situations above, there is a research gap in how to collect deep user feedback on a large scale.

In order to bridge the gap, personal reflection emerges as a prospective solution. For crowdsourced feedback to be effective, it needs to foster productive personal reflection on the preferences toward recommended items to generate useful ideas [49] for enhancing the performance of recommendations. By designing guided tasks and asking targeted questions, crowdsourcing tasks with personal reflection integration can motivate participants to reflect on their previous experiences and think more deeply about their choices, thus elaborating the reasons that lead to their preferences and generating profound insights. For instance, participants can be guided to reflect on whether the recommended items satisfy their needs, therefore providing a more detailed understanding. Additionally, personal reflection can facilitate participation and engagement by encouraging individuals to express and reflect on themselves. This approach contrasts with traditional micro-tasks and can somewhat mitigate the problems of existing crowdsourcing tools by not only prompting participants to delve deeper into their own data but also making them feel more valued to mitigate fatigue associated with repetitive tasks.

Realizing the possibilities of personal reflection naturally raises the question of how it can be effectively integrated into crowdsourcing tasks. To explore the potential of combining these two techniques, we propose a novel crowdsourcing method, seeking to fill the gap where crowdsourcing cannot meet the needs of recommendation systems by incorporating personal reflection technologies and methodologies into interactive task design. By enhancing user engagement through elaborately designed tasks, our solution addresses the limitations of a poor user experience and the insufficiently deep insights gained due to simple task design. Additionally, by leveraging the concept of personal reflection, participants are able to generate more profound insights than just simplistic thumb-up and thumb-down feedback. Finally, integrating personal reflection into the crowdsourcing paradigm, which utilizes collective intelligence, enables the aggregation and analysis of diverse perspectives from a large and varied participant pool.

1.1. Research Questions

In this study, the primary objective is to develop a crowdsourcing approach that incorporates personal reflection principles into task design. This approach aims to encourage the participants to reflect on their choices and motivations, ultimately providing deeper insights into the reasons behind their preferences and behaviors.

As such, the main research questions guiding this study are formulated as follows:

- **How to design personal reflection tasks in order to gather users' contextual factors and personal values for the YouTube recommendation system in the context of crowdsourc-**

ing?

- **What contextual factors and personal values for participants' usage of YouTube are extracted during one's reflection?**

1.2. Research Approach

This research begins with a literature study focusing on the existing user feedback mechanisms, crowdsourcing technologies, and personal reflection. This is done to build up a knowledge base to support the motivation of our research and consequently to present our approach (Chapter 2). In order to solve the first research question, Chapter 3 proceeds to the design and implementation of our approach, including the design principles it follows, the process of design, and the interface of the platform. Following the design, Chapter 4 conducts a user study involving 20 participants to evaluate the effectiveness of our proposed approach and presents using deductive coding to analyze the insights generated from the process, aiming to address the second research question. In Chapter 5, we reveal the analysis of contextual factors and personal values, respectively, the relationship between the co-occurrence of different values and specific contexts, together with a quantitative and qualitative analysis of the role of design principles in the reflection process. Chapter 6 discusses the implications of the insights and the lessons learned from the process of design, thus providing guidance for the YouTube recommendation systems and pointing to the directions for the design of similar studies in the future. Finally, in Chapter 7, we provide an overall summary of our research.

2
Related Work

2.1. User Feedback Mechanisms

In order to enhance the quality of their services, one notable trend in the field of recommendation systems is to rely on user feedback to gauge satisfaction levels and enhance the quality of the recommended content, therefore creating a need of collecting richer user experiences. User feedback is of vital importance as it provides valuable insights into user preferences, allowing recommendation systems to optimize their algorithms for a more personalized user experience. User feedback improvements can be effective in enhancing the performance of recommendation systems compared to purely algorithmic variants [4]. As a result, many feedback mechanisms and techniques within recommendation systems are proposed to improve the recommendations or suggestions for users, and they generally fall into two categories: explicit user feedback, such as ratings and comments, and implicit user feedback, like tracking user behavior and interaction patterns. Key user insights can be extracted from these feedback data to provide detailed information for customizing recommendations [37] and making more accurate suggestions tailored to individual user preferences.

2.1.1. Explicit User Feedback in Recommendation Systems

Explicit feedback, as its name suggests, could capture user preferences in a direct way [52]. It is more widely used, by explicitly asking users to comment to express their interests about a recommended item [37].

One common user feedback mechanism is the use of N meta-indicator rating systems, or rating an item using a scale. Typically, recommendation systems utilize an N-point Likert response scale. The scale's points are translated into numerical values that reflect the user preferences [36], and the ratings can help the systems make better future recommendations. In MovieLens, users can enter ratings through a five-star rating widget under the movie card to tell the system their preference for the movie [79]. Some systems, like YouTube or TiVo, also allow users to express their feedback through binary ratings such as thumbs-up or thumbs-down selection [69]. The interface of Twitter can also capture unary ratings (i.e., only positive ratings) such as the “Favorite” feature [37].

The numerical nature of the ratings allows for quantitative analysis, which can provide insights into user preferences, item popularity, and overall trends. By investigating the distributions of ratings and the trends over time, recommendation systems can be optimized to increase accuracy and relevance. However, despite the fact that mapping user feedback into a numerical scale is intuitive, it is not an appropriate way to measure user interests in the long run. In the first place, ratings have a certain level of uncertainty [15] due to the fact that user ratings inherently contain noise. This is because users may struggle to differentiate between movies they have viewed, even when re-rating experiments are conducted with a 1-day interval. Additionally, this approach requires users to perform an extra ”rating” action, which can be quite an overwhelming workload considering the enormous amount of digital content that users typically interact with on a daily basis [52]. This may result in the problem of data sparsity, where the sparsely populated rating matrix will ultimately limit the effectiveness of the recommendation. Finally, ratings restrict user expression and make it difficult to quantify the exact levels of user feedback. For example, a user rates a movie as 4 out of 5 stars can only mean that he thinks the film is generally well-made, but does not convey his opinions of whether he enjoys the plot, acting and other aspects.

In addition to ratings, there is literature on utilizing comments and product reviews [40] as explicit user feedback. User-generated content is also an additional source of knowledge for the recommendation systems. Various techniques, such as sentiment analysis, opinion summarization, and text classification, can be applied to process user comments as a source of indexing for product recommendations [16]. Wietsma et al. [73] proposed a recommendation system that incorporated user reviews to recommend hotels and attractions.

Utilizing user-generated content as explicit feedback for recommendation systems offers obvious benefits. It can provide better reasons for recommending products and increase user trust in the system. By analyzing user comments and reviews, recommendation systems can gain a deeper insight into the specific needs and interests of each user, resulting in more targeted recommendations. However, challenges also exist. In some areas, such as hotels and restaurants, changes in time and location may lead to shifting user preferences [30]. Additionally, biased or fake reviews could potentially mislead the systems to provide incorrect recommendations.

2.1.2. Implicit User Feedback in Recommendation Systems

Due to the fact that some users do not like to answer a lot of questions and provide explicit ratings, implicit user feedback increasingly becomes another useful source for analyzing user preferences. It is based on observable user behavior, such as user interactions, keyboard usage, and browsing patterns. For example, we show users the summarized reviews of several randomly selected rated products and then observe how they behave to implicitly infer user feedback [64]. There are a series of studies that investigate the relationship between implicit user feedback and their preferences. Konstan et al. [44] discovered that time spent reading correlated with ratings, where more time spent reading meant higher ratings. Hu et al. [33] proposed that in order to quantify users' interest in an item, we need to convert the implicit user feedback into user preferences approximately. Other forms of implicit user feedback include search history, watching history, mouse movements, and even eye tracking.

In addition to the advantage of convenient gathering and no requirement for extra effort from users, implicit user feedback can also improve the early-stage performance of a personalized system by reducing its uncertainty of users, especially when the users just start to use the system and have not provided much explicit feedback yet [81]. Nevertheless, inferring user preferences based on implicit feedback is not always appropriate. For instance, simply knowing that a user has purchased a cookie does not necessarily mean that he likes it, as the user may eat it before realizing that it tastes terrible. Similarly, staying on a web page for an extended period of time might not necessarily indicate interest; it could be the result of confusion or distraction. Besides, despite the fact that implicit action is more effective than explicit ratings when the purpose of the recommendation system is to increase engagement, the action-based system is not as accurate as the rating-based system, as it not only increases positive engagement but also raises negative engagement, like the negative action rate that correlates negatively with user satisfaction [78].

2.1.3. Summary

Based on the related work discussed above, it is evident that current research primarily focuses on exploring explicit and implicit user feedback mechanisms within recommendation systems. To sum up, explicit user feedback involves users providing direct input, like rating scales or comments. These mechanisms often rely on a numerical rating scale or N meta-indicator, making it more accurate. However, obtaining this feedback from users can be challenging, as it would be rather burdensome to enter a rating for every item, leading to low participation rates and biased data. Besides, a narrow focus on the numerical values of ratings probably limits users' expression and may not adequately capture and represent the exact nature of user preferences as well. Implicit user feedback, on the other hand, captures user behavior but may not fully represent the underlying reasons behind those behaviors. So it is easier to collect but less accurate in reflecting user preferences.

Despite the fact that these mechanisms play a crucial role in gauging user satisfaction and optimizing recommendation algorithms for a more personalized user experience, as the demand for personalization in recommendation systems continues to grow, it is clear that the data collected by applying these mechanisms is not able to meet the need of recommendation systems. Recommendation systems operate in a complicated environment where user preferences are influenced by various and nuanced factors, hence the research focus is supposed to be shifted to delve deeper into the intricate motivations behind user preferences and behaviors to gain richer user experiences. To effectively gather data, one promising method is utilizing crowdsourcing methodologies.

2.2. Existing Crowdsourcing Methodologies and Systems

Crowdsourcing emerges as an effective way to collect richer data. It allows for a vast amount of opinions and feedback from a variety of sources by harnessing the power of collective intelligence. In recent years, various crowdsourcing methodologies and platforms have been developed to gather data from user populations. Most of them involve direct participation and explicit collaboration from users to execute given problems. In the context of data collection, they are often classified into three main categories: voting system, information sharing system, creative system and game [75]. In this section, we are going to investigate the main categories of crowdsourcing systems and their applications in the field of data collection.

2.2.1. Voting Systems

In voting systems, crowd workers need to select their answers from several choices. Voting can be used as a tool to assess the correctness of the answers. Crowdsourcing marketplaces such as Amazon Mechanical Turk (MTurk) provide a vast number of experiments or tasks for participants to complete in an effective and flexible way online. Many tasks have a similar process and rationale as the voting systems.

- Evaluating tasks: Participants evaluate and rate ”items” (for example, books, movies, Web pages, and other users) using textual comments, numeric scores, or tags [46].
- Data annotation: Participants annotate or tag data, such as images, text, or videos, based on specific criteria or guidelines provided by the requester to create structured datasets. It has been proven that an accurate corpus can be built up by crowdsourcing workers [38], even with a lower error rate [21].
- Opinions: Crowdsourcing systems make it easy to gather opinions and subjective preferences from the crowd. Yang et al. [74] designed a basic type of human intelligence task for efficiently collecting user judgments on numerous dialogs, which used crowdsourcing through MTurk.
- Commonsense and reasoning tasks: Possession of commonsense knowledge and the ability to interpret are distinctive human capabilities. Many experiments are conducted in MTurk to collect commonsense knowledge and solve reasoning problems.

2.2.2. Information Sharing Systems

In information sharing systems, participants can easily share products, services, textual knowledge, and structured knowledge. For example, systems like YouTube can share products and services, such as videos. Twitter can share textual knowledge, and many science websites, like publication databases, can share structured knowledge. Aside from these types of information, some crowdsourcing systems can also share real-time research data. Some bike projects, such as CycleTrack [32], require users to record their trips by starting the app when they set out on a ride and then saving and uploading their data once they have reached their destination. The trip data is later used for planning facilities along the predicted routes. Another related example is Tiramisu Transit [82], which requires users to send real-time data on the vehicle, aiming to improve users’ transit experiences and transit accessibility. Moreover, Maisonneuve created the NoiseTube system, allowing users to measure their exposure to noise in their daily surroundings. The geo-location data and measurements can be automatically sent and shared online with the public, thus contributing to the development of urban noise maps.

2.2.3. Creative Systems

In creative systems, participants engage in tasks that require subjective input and creativity. These systems aim to take advantage of participants' unique perspectives and imagination, which cannot be replaced by any advanced technologies, to generate novel thoughts and create original content. They are valuable in fields like design, writing and painting, and sometimes they are conducted as creative contests to motivate more user support and participation. Through harnessing diverse insights from collective intelligence, creative systems have the potential to facilitate a breakthrough that traditional methodologies may not uncover. IdeasProject [45] is an open innovation and brainstorming community dedicated to harvesting ideas. Another example is Threadless [10], which is a platform for collecting graphic T-shirt designs.

2.2.4. Game

Humans might not be eager to complete some tasks unless they are presented in an appealing way. As a result, the games leverage people's desire to be entertained and produce useful metadata to solve problems efficiently with game players. Ahn et al. introduced Verbosity, a game to collect commonsense facts, which transformed the tedious process of entering facts in a database into an enjoyable game [72]. The online ESP Game [71] asks people to help determine the content of an image by providing meaningful labels for the image. To enhance games' effectiveness, researchers incorporate competition into them. There are several methods to motivate players, such as the use of high-score lists displaying the names and scores of the players who, within a certain period of time, achieve the highest number of points [70]. It provides strong and positive motivation for related tasks.

2.2.5. Advantages and Limitations of Current Crowdsourcing Systems

In the domain of recommendation systems, all of the four categories of crowdsourcing systems mentioned above can be applied to the development of personalized recommendations for users. Voting systems can collect valuable feedback and ratings on a recommended product provided by the participants to make analyses. By integrating voting systems into recommendation strategies, businesses can gather valuable insights from customers and enhance their products or services according to the feedback received. Information sharing systems involve users sharing their insightful opinions, unique experiences, and valuable perspectives with others. More knowledge resources may reduce participation costs and contribute to improved crowdsourcing performance [39]. Creative systems encourage participants to generate innovative ideas and content related to the recommended services, which can be used to provide more personalized recommendations to individual preferences. In games, tasks are packaged in an attractive and engaging form to create more enjoyable experiences or facilitate a strong willingness to participate.

Nevertheless, it is important to consider the potential challenges that come with utilizing crowdsourcing in recommendation systems. To begin with, tasks in voting systems may not fully satisfy the need to gather richer user reflections. This is because most of the voting systems are micro-tasks, like ratings, which typically involve providing brief comments or scores that lack depth and detail, as users may not have the opportunity or are reluctant to express nuanced opinions or elaborate on their experiences thoroughly. Additionally, these tasks typically involve one-time interactions, where users provide feedback on a specific item or question without the opportunity for further engagement or discussion. Lack of interaction and communication probably limit the depth of the collected insights. Secondly, information sharing systems may face issues related to data privacy and security. Users may be reluctant to share personal data due to concerns about whether their data will be potentially exposed. Thirdly, although creative systems can facilitate the generation of new ideas, however, without appropriate guidance and instructive task design, participants tend to produce irrelevant content, resulting in decreased efficiency. Last, task design presents a notable challenge in games. Tasks that are simplistic and boring, or those that demand excessive time to complete, will possibly lead to superficial user insights or discourage users from taking them seriously.

In summary, while existing crowdsourcing systems provide various approaches to gathering data, they have limitations that hinder their ability to fully meet the specific needs of recommendation systems to collect richer reflections on user preferences and behaviors. Current tasks often fall short of eliciting detailed and nuanced user feedback due to brief interactions or one-time engagements. Additionally, user engagement and willingness to participate can be hindered by privacy concerns and a lack of appeal. Apart from the shortcomings mentioned above, there is a scarcity of research in the overlapping fields of the two domains as most research focuses on micro-tasks. To bridge that gap and explore the possibility of leveraging crowdsourced data for recommendation improvement, there is a need to explore innovative methods. By integrating personal reflection into crowdsourcing methodologies, we can potentially address the problems in existing crowdsourcing methods to a certain extent, thus enhancing the quality and depth of insights collected.

2.3. Personal Reflection

To fix the issues with the current crowdsourcing techniques and investigate the exact reasons behind a user's choices and preferences, there has been a gradual focus of research on knowing oneself. One way to obtain self-knowledge is to reflect on one's personal information, such as behaviors, habitual patterns, and cognitive processes, namely personal reflection. Personal reflection refers to the process of knowing one's data that allows one to reflect on one's activities, make self-discoveries, and use that knowledge to make a contribution [12]. We propose the hypothesis that integrating personal reflection into traditional crowdsourcing platforms can enable users to reflect on their personal data, thus generating more nuanced and accurate insights into their preferences and behaviors. This can not only help to produce contextually richer and higher quality data compared to crowdsourcing micro-tasks, but also inspire users to elicit values [57] linked to their demand for recommendation systems. After collection, the specified data needs to be processed and transformed into various representations and different forms of information [51] for users to reflect on and analyze. At the stage of converting the data into reflective representations and motivating people to self-reflect, several methods and technologies have been proposed to facilitate this process.

Kurze et al. [47] presented ”Guess the Data”, a novel method where participants can conduct explorative work on their data and enable them to speculate and reflect on the data. During the process, participants are first required to interpret the data by guessing the origin of the data, identifying patterns, and making connections to everyday experiences. Then they engage in a collective discussion to share and reflect on their interpretations to generate insights into the data implications. Guessing establishes a link between participants' data and their everyday experiences, forcing them to recall the motivations and purposes behind their behaviors and consider whether their needs are satisfied, therefore generating insights and values about their preferences, habits and tendencies.

Another promising technology is visual exploration, which refers to the process of utilizing visualizations with personal data to help people reflect on their behaviors. Choe et al. [13] proposed an application ”Visualized Self” and found that data exploration with visualizations is a powerful way to assist individuals in uncovering meaningful insights about themselves and promoting self-reflection, as it motivates users to provoke questions and look for possible interpretations. To be specific, the special patterns and extreme values displayed in the visualization will probably catch people's attention, prompting them to reflect on what caused these unusual numbers. Pousman et al. [58] proposed Casual Information Visualization. Casual Infovis systems are designed to facilitate users' contemplation and reflection on their personal data by presenting it in a visually engaging and meaningful way. Through casual and accessible visualization, users are able to reflect on their experiences, memories, and emotions associated with the data, therefore deriving new perspectives from their digital traces. One example is Photomesa [7], where users can explore patterns in their photo collections to unearth unique insights.

As visualization emerges as an effective technique for self-reflection, research has been gradually focused on exploring the democratization of visualization [68], which is designed for non-experts to engage with data. Huron et al. [34] presented constructive visualization as a paradigm that provides users with the possibility to create simple and flexible visualizations. For example, shifting the focus range to choose and compare various time spans enables users to elicit memories of prior behaviors and evoke external contexts [13]. This approach empowers users to engage with their data in an intuitive way and delve deeper into their behaviors during the process of interaction. By constructing visual representations of their own, they can probably spot patterns and connections that were previously hidden. In the context of crowdsourcing, compared to micro-tasks, this self-motivated exploration encourages users to be more involved in tasks and provide richer feedback.

Lifelogging technologies [11], especially storytelling also appear to be a useful method for introspection.

The act of recalling and reliving past events in the form of a story provides a medium for the exchange of individual experiences. This narrative form of introspection allows users to reflect on themselves and gain a deeper understanding through the process of storytelling by identifying significant moments and changes in their past lives and making sense of their experiences in a structured manner. An effective form of storytelling is user journey mapping, which is a useful way to construct and visualize users' experiences [3]. It enables the individuals to plot out the key events and the experiences associated with them. By visualizing their story, users can recall important activities, spot special patterns, and recognize long-term trends. Users are able to directly realize the evolution of their desires and experiences over time and provide comprehensive narratives, from which we can analyze and extract insights.

Overall, these various methods and technologies contribute to a growing body of research aimed at leveraging personal reflection to better understand humans. As such, integrating personal reflection technologies into crowdsourcing platforms may offer a possible solution to bridge the gap between the fields of crowdsourcing and recommendation systems and satisfy the latter's need. By combining the strengths of these two technologies, it is possible to compensate for the shortcomings of crowdsourcing to some extent and ultimately reach the goal of collecting richer insights comprehensively and effectively. In response to this hypothesis, we propose a novel method that incorporates personal reflection technologies into crowdsourcing methodologies, which naturally leads to our first sub-research question: **How to design personal reflection tasks in order to gather users' contextual factors and personal values of their preferences and behaviors?**

3 Design and Implementation

The objective of this chapter is to address the first research question:

- *How to design personal reflection tasks in order to gather users' contextual factors and personal values for the YouTube recommendation system in the context of crowdsourcing?*

In this chapter, we will delve into the design and implementation aspects of our project. This chapter is structured into three sections: design principles, design process, and implementation. We will begin by discussing the various considerations and principles that influenced our design decisions. Following this, we will outline the iterations of our design, detailing the steps of modifications and the rationale in each phase. Finally, we will provide a comprehensive explanation of the implementation stage, including the development techniques and methodologies utilized to realize the design.

3.1. Design Principles

In designing the task of our project, several key principles guided our approach to ensure that participants could effectively reflect on their YouTube-watching history, generate insights about their preferences and behaviors, and elicit their personal values regarding the recommendation system.

DP1 **Fully Reflect by Slowing Down the Thinking Process**. To facilitate deep reflection, this design principle emphasizes the importance of slowing down the thinking process. It is grounded in the idea that reflection requires time, by encouraging participants to take time to consider the reasons behind their preferences and behaviors, they may be able to connect the data with their personal experiences. According to prior research, slow technology [23] is a promising solution to prompt reflection and moments of mental rest. Through deliberate user interfaces, subtle interactions, and components that encourage users to slow down and contemplate, users are more likely to empathize and become immersed, eliciting insightful reflections. One effective method to achieve this principle is to guide participants through a process of speculation about their past behaviors followed by verification against actual data. The speculation involves asking participants to take time to recall their preferences and habits to make connections to their data [47]. The comparison is a reflective process, as it forces the participants to slow down and carefully examine the similarities and differences between their expectations and real-life situations to make explanations for special phenomena. During the slow thinking process, participants are able to reconsider and reconstruct their understanding of their behaviors, ultimately gaining more comprehensive and profound reflections.

DP2 **Contextual Reflection and Judgement**. The task in our project should encourage the participants to reflect on their viewing habits and make connections to their personal values within the context of personal circumstances and daily lives. This design principle is implemented by prompting participants to recall the specific situations and purposes behind their video-watching behaviors and evaluate the videos within their context. Prior literature revealed that incorporating context into the thinking process can facilitate reflection by making relevant connections to meaning within the context [19]. Additionally, There is also literature indicating that evaluative judgments prove to be effective in helping people elicit personal values [20]. As such, our platform needs to empower participants to recall and describe the specific contexts in which they watched the video, such as their emotional state, the special events, or the purposes, followed by an evaluative judgment on their viewing habits, allowing them to connect their viewing behaviors with their personal experiences and routines and generate more meaningful insights.

DP3 **Guidance-Based Reflection**. The task in our project should promote progressively deeper reflection by providing structured guidance through the process. This principle is implemented by guiding participants from an overview of their data to a more detailed analysis step by step through the whole process of the task, together with instructive questions in each section to enable participants to progressively dig into their data. Many prior works suggested a need for interventions and encouragement to support reflection [6]. Moreover, according to the previous paper, providing justifications or explanations for one's actions, decisions, and experiences can facilitate personal reflection [18]. Therefore, providing specific guided reflection questions can help to guide their thinking and reflection [19]. As a result, our platform is supposed to design a step-by-step deepening process to encourage participants to gradually connect with their data by moving from overall patterns to specific spots. Furthermore, in each section, by incorporating tailored instructions and well-structured questions, participants are able to start from broad observations, and then to specific interpretations behind the patterns by making connections to life context, ultimately eliciting values by making evaluative judgments.

3.2. Design Iterations

The design of our task evolved through several iterations, each influenced by usability testing and user feedback. This iterative process was aimed at not only refining user experience and functionality but also being aligned with our objectives of encouraging users to reflect on their data and generate reflective insights.

In the design iterations, we initially set up three sections. The first section involved requiring participants to guess their most-watched video types and time periods and then compare their guesses with the actual data presented in straightforward charts. This section proved to be proved to be feasible since it effectively enabled participants to discover how their expectations aligned with reality, prompting them to make interpretations and initially reflect on their viewing habits.

The second section began with a design that asked participants to set colors for different time periods and design icons for each video category. These elements were then filled into the corresponding date in the given calendar. We aimed to encourage deeper engagement and make the reflection process more intuitive by allowing individuals to design custom visuals to represent their data [42] and adding some game design qualities [54]. However, this version included an excessive workload due to the time-consuming and repetitive work of designing and dragging icons, which made little contribution to facilitating personal reflection. In response to this feedback, we shifted to investigating the effects of data reconstruction. According to prior literature, The paradigm of constructive visualization [34] was promising in facilitating self-reflection as it encouraged proactive interaction with the data [66]. As such, we provided participants with various angles to help them explore the relationships and special patterns from the combinations based on the number of videos, video categories, time periods, and days of the week. Yet, after a few pilots, this version appeared to contain too many angles, leading to users feeling overwhelmed and confused, and a large proportion of participants' answers from different angles were overlapping. In this case, we simplified the second section by removing some perspectives and ended up with just two. This iteration proved to be more successful as it not only improved user experience but also collected sufficient data.

The third section was originally designed to recall the emotions associated with each video and put these emotions on the timeline. We created the timeline with the intention of displaying one's emotional style to foster self-reflection [77]. Nevertheless, this version proved problematic because users found it difficult to recall their emotions for each video. Additionally, the categorization of emotions was overly complex, resulting in an unclear timeline. Consequently, we revised this section by focusing on a specific day from last week and asking users to select a few videos and create a story map to demonstrate mood changes. However, the feedback indicated that it was hard for users to recall emotions accurately, therefore influencing their reflection on previous behaviors. Accordingly, we turned to analyzing users' evaluation of the video by instructing them to recall its usage in their living scenario. This version turned out to be more effective as we encouraged users to think about how videos fit into their daily lives and how they fulfill their needs or purposes at the time.

In addition to specific task design, we iterated other aspects of our platform as well. We initially set up all the sections in the Miro board, enabling participants to interact with operable objects, but most of them were unfamiliar with Miro, which greatly reduced the efficiency of our approach. Thus, we displayed the content of the first two sections directly on the web page and left only the last section in the Miro. Furthermore, we improved the reflective questions in each section to make them more thought-provoking and modified other text instructions to avoid situations where participants were confused. We also beautified the layout of the web pages to enhance the user experience further.

3.3. Implementation

By incorporating the design principles and the back-end implementation, we established a prototype that effectively demonstrates the functionality of our platform. This section details the implementation of both the user interface and the back-end, as well as the deployment of our platform. The code repository is now available online [1], and the software environment is shown in Table 3.1.

Environment Setting	Parameters
Programming Languages	JavaScript
Development IDE Setup	Visual Studio Code + Volar (and disable Vetur) + TypeScript Vue Plugin (Volar)
Environment Configuration	Vue.js, Node.js, Nginx

Table 3.1: Software Environment

3.3.1. User Interface

To facilitate the interactions between participants and their history data to prompt personal reflection, the user interface(UI) of our platform was developed using Vue 3, focusing on creating a responsive, intuitive, and user-friendly experience.

The UI implementation contained several functional modules and technology stacks. Components were developed using Vue.js's component-based architecture to encapsulate HTML, CSS, and JavaScript within a single file. Integrated Vuex was used for state management to ensure that data shared across multiple components remained consistent. Vue Router handled navigation and routing, allowing for smooth view transitions. Additionally, API integration, like Axios, allowed the front end to send HTTP requests to communicate with the back-end server, enabling dynamic content and real-time updates.

In addition to these libraries that ensured a cohesive and effective UI, there were several main sections that integrated the aforementioned design principles, making our platform strike a balance between functionality and user experience. Our method was designed in three sections, with DP3 running throughout. The progressively deepening process started from a broad overview of their viewing habits to a deeper reflection on the analysis of specific videos (DP3: Guidance-Based Reflection). During the whole process, we also proposed multiple reflective and thought-provoking questions in each section (DP3: Guidance-Based Reflection) to help participants reflect step by step, from a simple description of data, then explaining the reasons behind the patterns within the life context and finally evaluate their watching behaviors to elicit values for the YouTube recommendation system. The three sections are described below.

Section 1. Make a Guess About Your YouTube Watching Habits

This section enables participants to gain an initial understanding of their viewing habits by making simple speculations (DP1: Fully Reflect by Slowing Down the Thinking Process), then explaining their

[1] https://github.com/Desmond766/Insights-investigation/tree/final_yizhen

guess within the life context and comparing their expectations to actual data, and finally connecting their watching habits to their values (DP2: Contextual Reflection and Judgement). We designed and displayed several reflective questions and visualizations to assist participants in engaging in a process of self-reflection(DP3: Guidance-Based Reflection). This section is divided into two parts: the speculation part and the verification part. The speculation part comprises two questions, each divided into two sub-questions, requiring participants to guess the most-watched video categories and time periods over the last month and provide corresponding reasons for their guesses, respectively. A preliminary guess enables them to take the time to recall their viewing habits, thus getting a first look at themselves and making preparations for the next step of reflection. Interpreting the reasons for such speculation empowers participants to connect to their daily lives, helping them to explain their behaviors within life circumstances.

To facilitate the process of self-reflection, in the verification part, we displayed the bar charts of the number of the top four categories of videos and the number of videos they watched in different time periods. Participants then compare these visualizations with their initial guesses and make explanations for the results of the comparison following our instructive questions. This process motivates participants to learn about themselves from the difference between the guess and the actual data through personal reflection during the process of making interpretations. At last, participants are required to reflect on whether their watching habits align with their personal goals to facilitate the elicitation of their values for the YouTube recommendation system.

Figure 3.1 and 3.2 illustrate how participants interact with our platform in Section 1. Details of the questions are in Table 3.2. In Step 1 of Section 1, they began by making speculations about their viewing habits. The category of videos refers to the classification of video content, like sports, music, and study. The time periods are predefined, dividing the day into six parts corresponding to bedtime, morning, lunchtime, afternoon, dinnertime, and evening. Next, they are required to explain the reasons for their guess in the context of their personal habits and circumstances. In Step 3, participants are displayed two bar charts generated by their watch history. Then they are asked to compare their guesses to the visualizations and reflect on their data. If the guesses align, they are required to explain how viewing habits align with their personal goals and values. If the guesses do not align, participants are prompted to reconstruct their understandings of themselves.

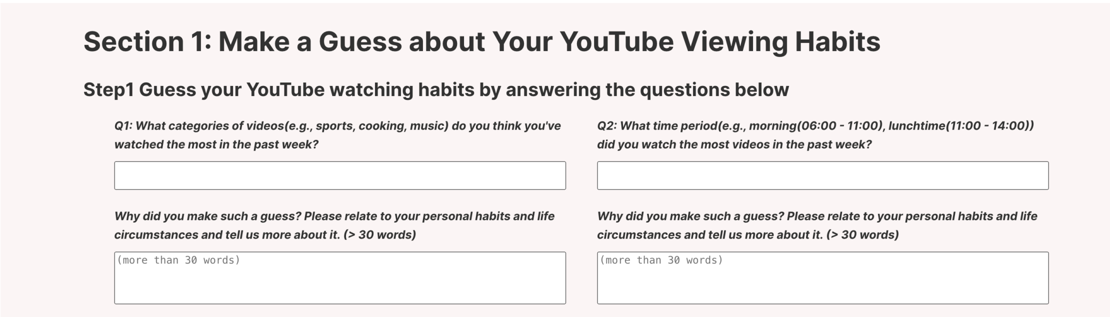

Figure 3.1: Interfaces for Section 1, Step 1

Question Number	Question
Q1	What categories of videos(e.g., sports, cooking, music) do you think you've watched the most in the past week? Why did you make such a guess? Please relate to your personal habits and life circumstances and tell us more about it. (> 30 words)

Question Number	Question
Q2	What time period(e.g., sleeping time(00:00 - 06:00), morning(06:00 - 11:00), lunchtime(11:00 - 14:00), afternoon(14:00-17:00), dinnertime(17:00-20:00), evening(20:00 - 24:00)) did you watch the most videos in the past week? Why did you make such a guess? Please relate to your personal habits and life circumstances and tell us more about it. (> 30 words)
Q3	Reflecting on Figure 1, how does your guess of the top category of videos align with the actual data (e.g., the top category of video and its number)? (> 20 words)
Q4	If your guess aligns with the actual data, how does this watching habit align with your personal goal and value? And why? If your guess is different from the actual data, what did you learn about yourself? (> 30 words)
Q5	Reflecting on Figure 1, how does your guess of the top category of videos align with the actual data (e.g., the top category of video and its number)? (> 20 words)
Q6	If your guess aligns with the actual data, how does this watching habit align with your personal goal and value? And why? If your guess is different from the actual data, what did you learn about yourself? (> 30 words)

Table 3.2: Questions in Section 1

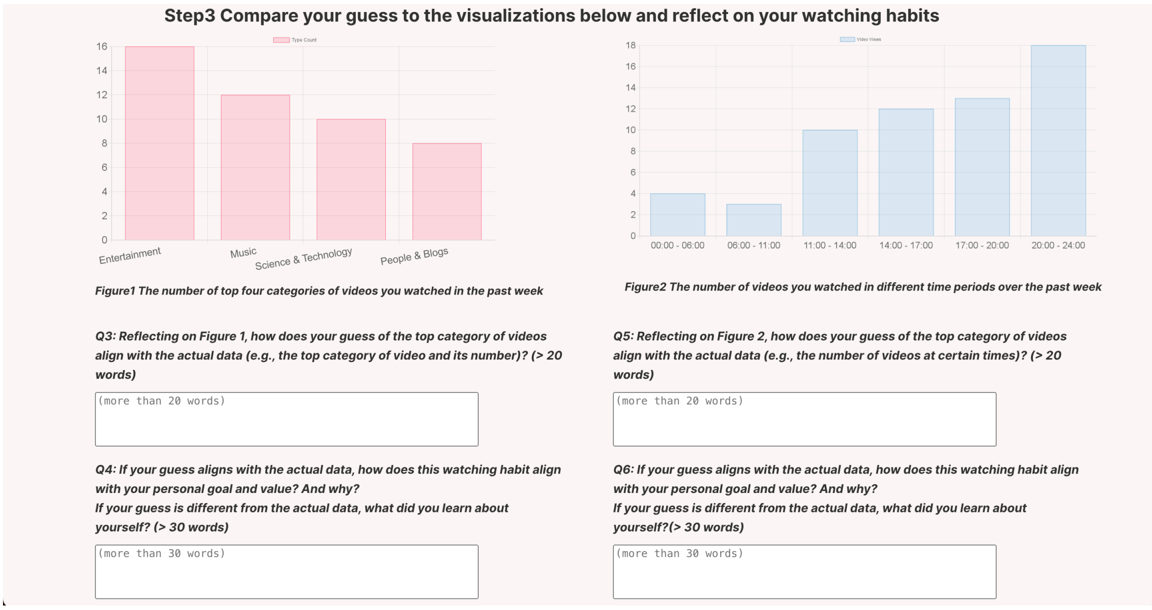

Figure 3.2: Interfaces for Section 1, Step 3

Section 2. Reflecting on YouTube watching Purpose through Video Category or Time Period

In this section, we require the participants to choose one of the two perspectives, corresponding to the two perspectives of reflecting on viewing habits in the previous section, i.e. video category and time period, based on the results of the comparisons in Section 1, to further analyze their viewing habits from a specific angle of interest.

We provide participants with two perspectives, and they can switch between them to carefully consider which angle is more meaningful (DP1: Fully Reflect by Slowing Down the Thinking Process). The first perspective focuses on looking into the distribution of video categories across different days of the week, and the second perspective concentrates on exploring the number of different categories of videos participants watched in each time period. In each perspective, we set well-structured reflective questions to progressively deepen their self-reflection (DP3: Guidance-Based Reflection), ranging from describing the phenomena, like special watching patterns, to interpreting the reasons behind them within the context of habits and life circumstances and finally linking them to the individual's own personal values by evaluating their watching habits (DP2: Contextual Reflection and Judgement).

Figure 3.3 and 3.4 show the detailed process of Section 2. Two angle buttons are displayed under the instructions. Participants are prompted to choose either to further reflect upon their YouTube-watching behaviors. In both perspectives, there are corresponding charts to help participants analyze their viewing patterns. In the angle regarding to video category, we display a stacked chart showing the distribution of the video categories across different days in the last week. Participants can observe how their preferences for specific video categories fluctuate throughout the week. In the time period angle, participants are shown a stacked chart illustrating the distribution of different video categories in each time period. Participants can reflect on the range of video categories they watch during different time periods throughout the day.

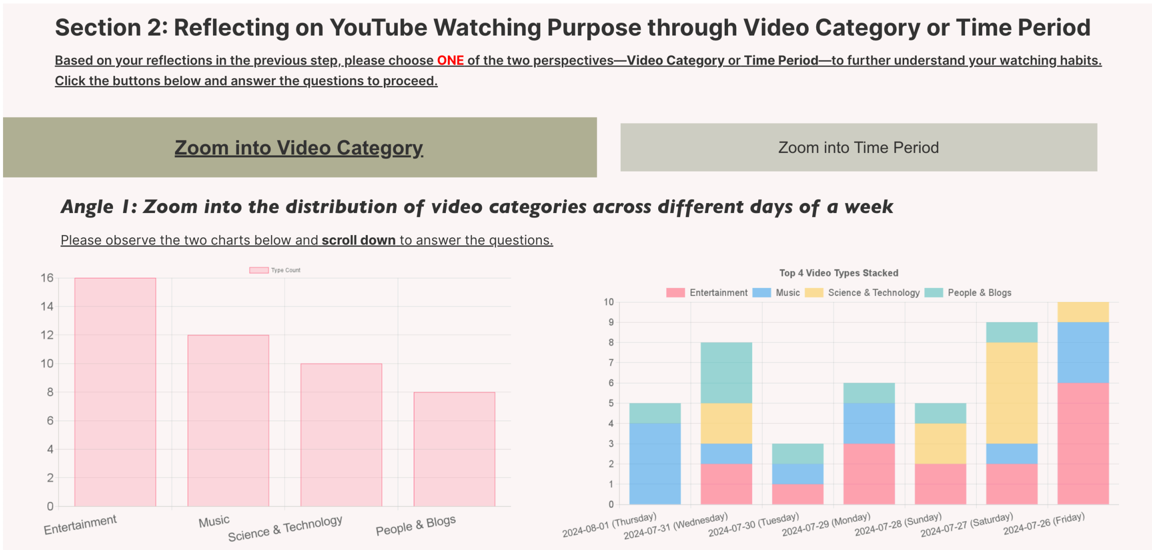

Figure 3.3: Interfaces for Section 2, Angle 1

Both sections contain three reflective questions, which are shown in Table 3.3. Participants begin by looking for interesting patterns in their viewing habits. For example, on weekends, they may discover a significant increase in the number of entertainment-related videos, while morning routines may reveal a tendency to news-related content. Subsequently, participants need to explain the patterns in relation to their habits and life context, such as a demand for relaxation after a busy week. Finally, we instruct participants to understand how their patterns align with their personal values by encouraging them to reflect on whether their viewing habits can satisfy them or not.

Question Number	Question
Q1	Do you notice any interesting patterns in the charts above? For example, do you observe certain categories of videos are watched more on certain days or differences between weekdays and weekends? (> 20 words)
Q2	What factors contribute to this observed pattern? Please relate to your personal habits and life circumstances and tell us more about it. (> 40 words)

Question Number	Question
Q3	How are you satisfied with this watching habit represented by the data pattern? Please relate it to your personal values and tell me more about it. (> 40 words)
Q4	Do you notice any interesting patterns in the chart? For example, did you tend to watch a certain category of videos at a certain time period, or did the number of views vary between time periods? (> 20 words)
Q5	What is the reason behind this pattern? Please relate to your personal habits and life circumstances to explain more about why you choose those videos. (> 40 words)
Q6	How are you satisfied with this watching habit represented by the data pattern? Please relate it to your personal values and tell me more about it. (> 40 words)

Table 3.3: Questions in Section 2

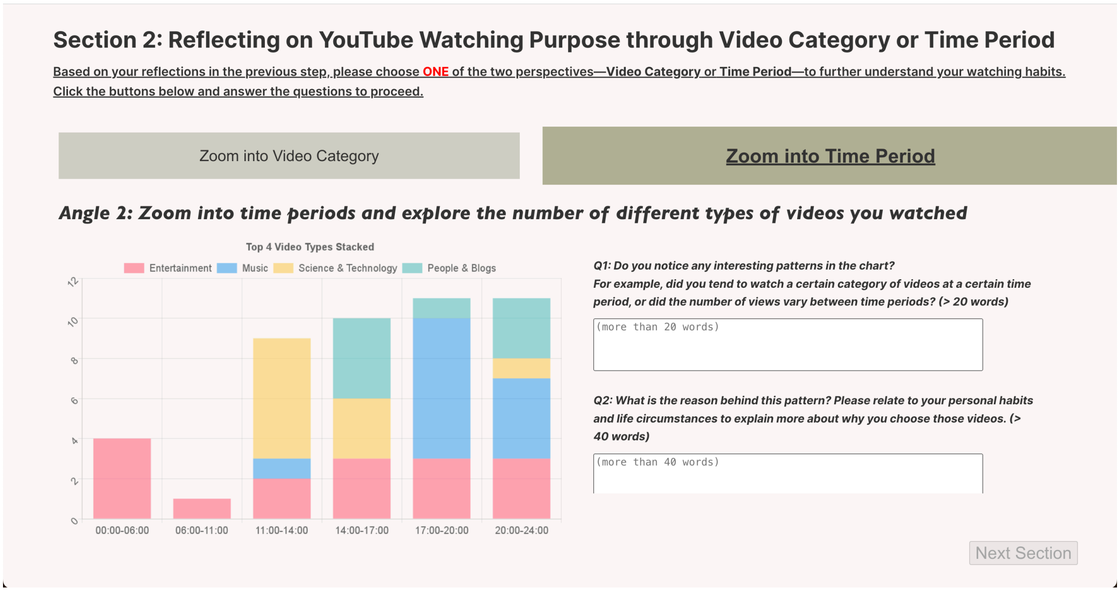

Figure 3.4: Interfaces for Section 2, Angle 2

Section 3. Evaluate the YouTube video by recalling its usage in your living scenario

In this section, we designed a task that focused on evaluating YouTube videos based on their usage in specific life contexts to help participants self-reflect and elicit personal values of YouTube videos. This section involves asking the user to select a day from the previous week and then guiding them to create a story map of life activities from this specific day and evaluate the videos they have watched. This is a further in-depth exploration of the previous section, as we move from an analysis of the weekly pattern to a more specific focus on a certain day (DP3: Guidance-Based Reflection). We provide participants with an opportunity to create their own story map to help them become more engaged and connect their experiences to the videos, enabling them to place themselves in context in their subsequent reflections (DP1: Fully Reflect by Slowing Down the Thinking Process). By evaluating their satisfaction level with each video in the context of their activities, participants can make connections between their personal experiences and values, thus reflecting contextually on their behaviors and eliciting their values for recommendation systems (DP2: Contextual Reflection and Judgement).

Figure 3.5 demonstrates the interactive interface for the step of creating a story map. The interface is divided into two parts. On the right side of the page, there is a Miro board, where participants need to follow the instructions to construct their own story map. The detailed content of the Miro board is shown in Figure 3.6. We generate a timeline inside the coordinate axis showing the videos they watched through the day they selected, and a reference table showing the information of each video, including title, author, watching time, and category. Participants start by reconstructing their day by recalling the activities of that day according to the timeline. They then select five videos from the timeline by referring to the reference table and evaluate the satisfaction level of each video based on their viewing purposes.

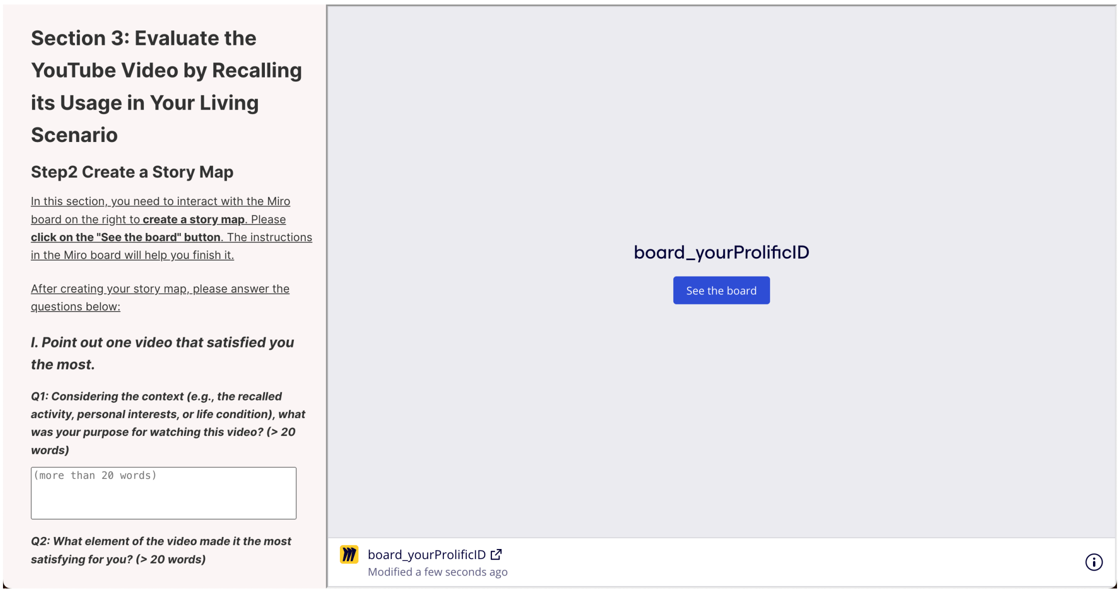

Figure 3.5: Interfaces for Section 3

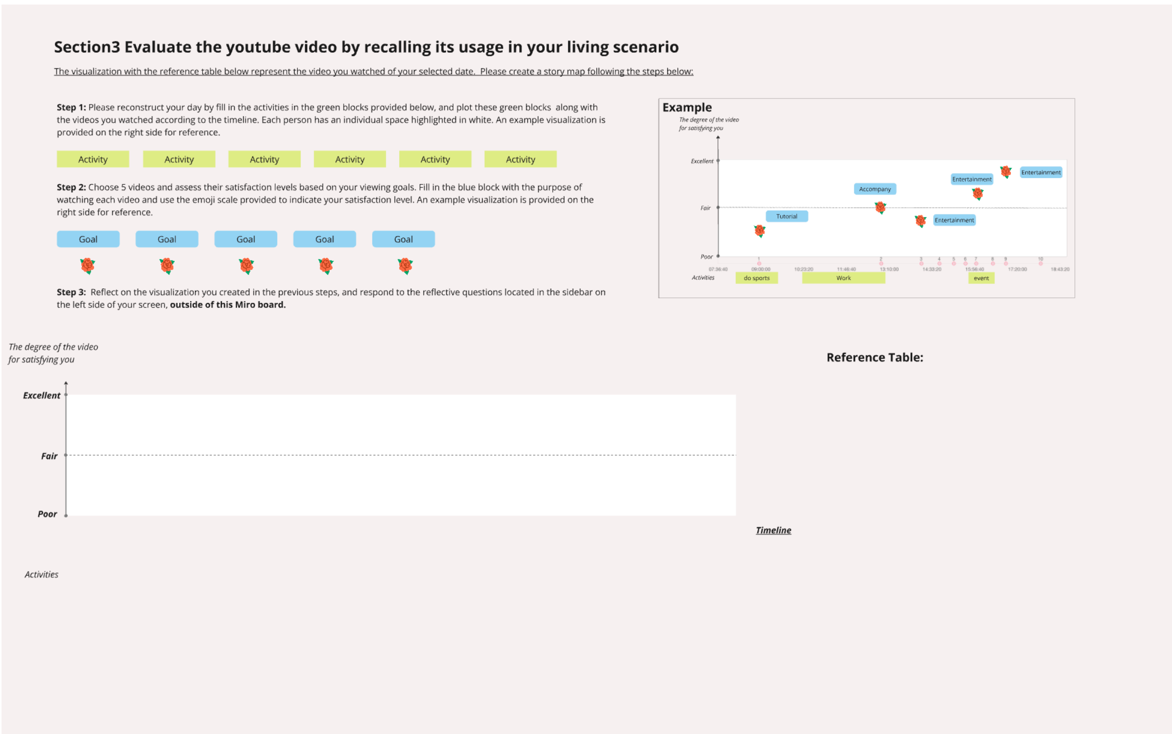

Figure 3.6: Content of the Miro Board

After creating their story map, they are required to answer the reflective questions on the left side of the page. Details of the questions are shown in Table 3.4. Based on the visualizations they created, they need to reflect on their most and least satisfying videos respectively. For each video, participants are guided to first explain the reasons why they chose to watch this video within the context of the recalled activities or personal preferences. After that, they consider what elements of this video make it most/least satisfying for them, followed by interpreting the reasons and elaborating the connections to their personal values.

Question Number	Question
Point out one video that satisfied you the most.	
Q1	Considering the context (e.g., the recalled activity, personal interests, or life condition), what is your purpose for watching this video? (> 20 words)
Q2	What element of the video makes it the most satisfying for you? (> 20 words)
Q3	Why are you satisfied by this particular element? How does it relate to your personal values? (> 40 words)
Point out one video that satisfied you the least.	
Q4	Considering the context (e.g., the recalled activity, personal interests, or life condition), what is your purpose for watching this video? (> 20 words)
Q5	What elements of the video made you feel less satisfied? (> 20 words)
Q6	Why do you dislike these elements? How do they against your personal values? Tell us more about it. (> 40 words)

Table 3.4: Questions in Section 3

3.3.2. Back-end Implementation

The back-end was implemented using Node.js as the server-side operating environment and Express.js as the web application framework with various middleware integrations, providing a robust server-side infrastructure. The back end handled various tasks, as outlined in the following:

Real-time communication was a key function, implemented utilizing Socket.io, which enabled real-time updates and interactive sessions. Socket.io was of vital importance to the workflow of our platform, as it acted as a communication bridge between the client side and the server side. The interactions on the client side would trigger the server-side processes such as copying the Miro board and processing the uploaded YouTube-watching data.

The uploads of all files, including YouTube-watching JSON files, TXT files generated from users' answers, etc., were managed by Multer, and the server defined various API endpoints to handle these requests. After appropriate processing, the JSON data in the uploaded YouTube watching history could be parsed for further operations. For example, using YouTube Data API v3 to obtain the category of the video from its URL. Another instance is generating visualizations. Chart.js and ChartJSNodeCanvas were utilized for server-side chart rendering, enabling the generation of these visualizations. The rendered charts were then displayed on the front end or used to update the Miro board, providing users with interactive data representations.

3.3.3. Deployment

Our project was deployed to a server using Nginx, which was an open-source high-performance web server and reverse proxy [60] that could handle client requests effectively. Nginx not only supports load balancing, which allows requests to be distributed to multiple back-end servers, but it also excels at handling highly concurrent requests, making it an ideal choice for the deployment of our project.

After installing and starting Nginx, we packaged the project and converted the source code to static files. The Nginx configuration file was required to be edited to specify: 1) the virtual host to handle client requests; 2) the listening port. Listening on port 443 requires an SSL certificate to provide HTTPS services; 3) location blocks, which are defined to handle different types of requests by specifying the URL path of the request and the corresponding processing method; 4) root, which refers to the packaging output directory of our project. Finally, adjust the firewall settings to allow traffic on the specified port (typically 80 or 443 for HTTP and HTTPS, respectively).

4
User Study

Based on the design considerations proposed in the previous chapter, we introduced a design framework that incorporates several personal reflection techniques. Then we implemented it in the crowdsourcing paradigm, thus generating a prototype.

In order to evaluate the effectiveness and performance of the prototype, we designed and conducted a user study to see if this approach can collect deep user feedback in a crowdsourcing context and analyze the extracted contextual factors, personal values, and design principles. The experiment was conducted in July and August 2024. The study was conducted on the Prolific Online Platform, an essential tool that allows researchers to run scientific research and find vetted research participants.

We start by providing a brief introduction of the context and required materials, followed by detailed information about participants, procedures, and the statistical analysis techniques utilized for result analysis.

4.1. Context and Materials

We conducted this study in the context of people using YouTube to watch videos for several reasons. First, YouTube serves as a source of entertainment, education, and social connection for millions of individuals worldwide. It is widely used, therefore we are able to collect sufficient watch history on a global scale and from distinguished backgrounds for further analysis. Second, many people use YouTube to watch videos on a daily basis, which will shape their watching preferences and behaviors. Prior work has indicated that users' behaviors of watching videos are influenced not only by situational characteristics [43], referring to the context of interaction [53], but also by personal characteristics, indicating individuals' values and demands for the recommendation systems. Additionally, research has shown that engaging in a reflection on one's previous behaviors can help one gain a deep understanding of their life patterns, self-identity, and the significance of this past experience to their sense of self [11].

In order to analyze users' contextual factors and personal values behind their behaviors when using YouTube, we collected the YouTube watch history from each participant for further processing. The YouTube watch history includes the title, author, URL, and watch time of the video. We also utilized YouTube API 3 to extract the category of the video. Besides this, We also utilized other several materials to ensure that the experiment went smoothly.

- **Informed Consent Form**: Participants were first provided with an informed consent form detailing the purpose and potential risks of our study and the intended use of their data. Participants can only join in the experiment after they sign the consent form.
- **Text Instructions**: During the process of our study, participants were required to follow the text instructions to complete all the sections. The text instructions contained a tutorial for downloading YouTube watching history and explanations guiding participants to reflect on their data.
- **Prototype**: The prototype was designed and implemented as described in Chapter 3.

4.2. Participants

We recruited participants through the Prolific platform, which provides access to a diverse pool of pre-screened individuals. The inclusion criteria for participants in our study were as follows:

- Age: Participants had to be 18 years of age or older.
- Language: Participants needed to be fluent in English as they are required to understand and respond to the study materials that are in English.
- Prolific Approval Rate: Only participants with an approval rate higher than 95% on previous studies were eligible for our study to ensure the quality of the data.
- Social Media: Only participants who used YouTube regularly in their daily lives were eligible for our study since we aimed to investigate the reasons behind the YouTube videos they had watched.

We recruited a total of 20 people from the eligible participants screened based on these criteria. All of them managed to sign the consent form and complete the study. Participant privacy was protected by anonymizing all data collected in the study, as the focus was on gathering user reflections without recording identities. We only had access to their Prolific IDs, which were used to match their answers and ensure they completed the study. Each participant managed to complete the study successfully, and all of their data was included in the analysis.

4.3. Study Setup and Procedure

The procedure of our study for any participant is outlined in Figure 4.1. After we published the study on the Prolific Platform, those interested were able to participate and they would then be redirected to our platform URL. Each participant took part in a scheduled 50-minute study to reflect on their YouTube watching history. Before the participants formally entered the task page, we executed several preparation steps. Participants were first given a brief explanation of the content and purpose of this study and informed of the need to collect their YouTube viewing history data as well as screen recordings, and then they were required to sign the consent form. Following that, they were shown detailed instructions on how to download their YouTube history. The history data was downloaded from Google Takeout in the form of a JSON file. Since the download would take a few minutes, participants could go to the next step first, open the screen recording, and enter their Prolific ID, enabling us to pair them with their answers. After exporting the file, participants needed to upload the JSON file, which we processed for subsequent operations, such as generating visualizations.

Once everything was in place, the participants were required to complete all the sections of the task without further assistance from the researcher. The task contains three phases:

- **Guess Watching Habits Phase**: Participants self-assessed and predicted their YouTube watching habits by answering some questions, and then they compared their guesses to visualizations generated from their true data. This section allowed participants to explore their viewing patterns and gain insights into their behaviors.
- **Analyze Viewing Patterns Phase**: Participants were provided with two perspectives, with which they were able to choose one to deepen their understanding of their YouTube-watching behaviors, and answer several questions. By selecting the angle of interest, participants could zoom into the distribution of the top four video categories across the days in the week or within different time periods to discover special patterns, explore the reasons behind them, and align the patterns to their values.
- **Recall and Evaluate Phase**: By creating a story map of life activities from a specific day in the Miro board, participants were guided to recall the usage of YouTube videos they watched in their living scenario, and then be asked to answer a couple of questions to evaluate the videos within their living scenario and elicit their values about recommended videos.

After participants completed all the sections, we provided them with a completion path, from which they were able to be redirected to Prolific, and it would show in our experiment interface that they had submitted the correct completion code. The submissions were therefore reviewed manually by us. Both the correct completion code and the quality of submissions are taken into consideration to deciding on payment. During the whole process, participants answered the questions and interacted with the screen following the instructions. All answers and their interactions with the Miro board were recorded with their knowledge.

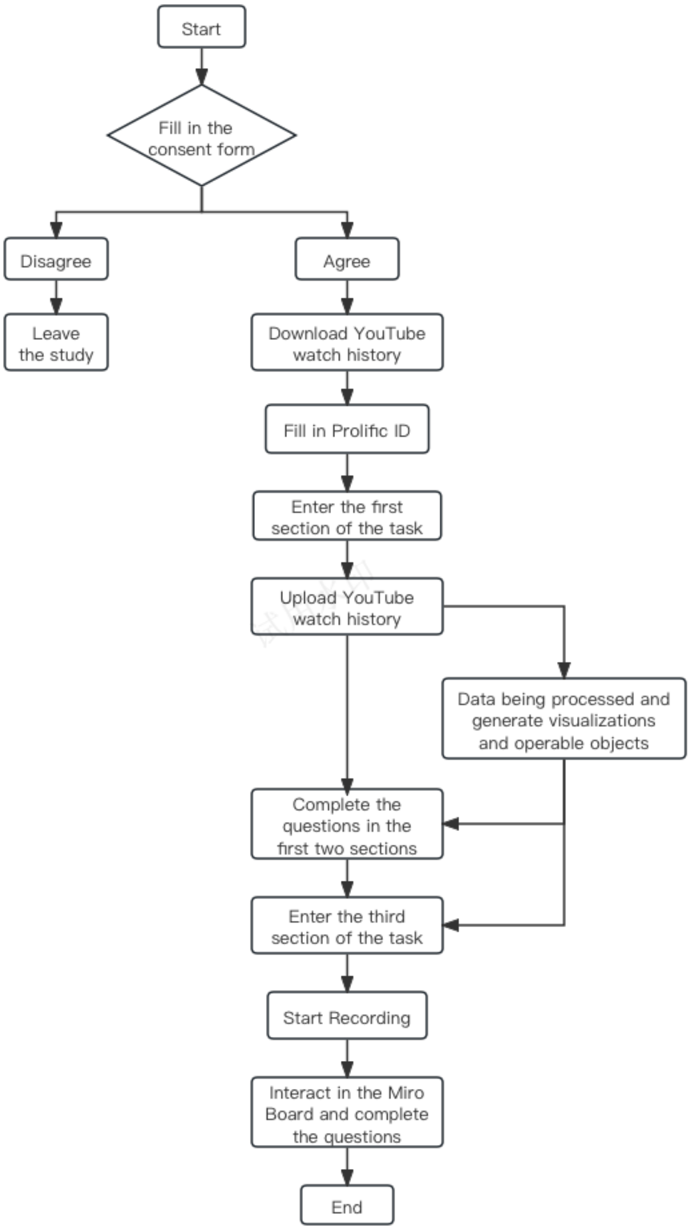

Figure 4.1: Pipeline of the Proposed Platform

4.4. Approaches to Data Analysis

We aim to gain a qualitative understanding of the process of reflection, the recalled contextual factors, and elicited personal values. As such, we captured each participant's answers to all reflective questions. We analyzed all the answer files to understand and evaluate the generation of insights and whether the design principles we proposed had worked successfully in the reflection process.

To draw conclusions and gain insights into participants' perspectives, researchers need to annotate each data item with a code, a process known as coding [26]. In our research, we utilized deductive coding. It refers to the process of first defining an initial organizing framework comprising of themes or codebook [5] and then applying it to the new participants' answers as the analysis of the study [17].

Our choice of software was *Atlas.ti*, a powerful workbench for qualitative data analysis that enables interpreting texts with coding [63]. It also provides functions of analysis within codes and files.

For the 20 groups, we needed to code manually. We began with discussing and establishing initial codebooks for contextual factors, personal values, and design principles, respectively.

Contextual Characteristics are defined as information that describes the situation and the environment of the individuals involved in the recommendation system [1]. In our project, we started by coding all the contextual factors and then merged and classified the codes into several categories, namely task context, temporal context, social context, cognitive context, physical context, and application context [48].

Personal values are defined as what a person considers important in life [9]. Incorporating human values into recommendation systems can ensure an ethical and user-centered environment that benefits both individuals and societies [65]. We therefore focused on a couple of themes of values that might be important to consider in recommendation systems based on previous work, such as *Well-being*, and *Usefulness*; each is further subdivided into several types.

For design principles, we further divided the main principles mentioned in Chapter 3 into several sub-codes. DP1, **Fully Reflect by Slowing Down the Thinking Process**, is divided into three more detailed directions:

- *Speculation and Comparison*: This sub-code suggests encouraging users to actively engage in the process of guessing their watching habits and comparing them to the actual data. As such, this sub-code corresponds to Q1, Q2, Q3, and Q5 in Section 1.
- *Different Angles*: This sub-code highlights encouraging users to choose a perspective that is more interesting or that they find special patterns to reflect on. It corresponds to the answers to Q1 and Q4 in Section 2.
- *Constructing a Story Map*: This sub-code emphasizes deeper reflection by motivating participants to narrate their activities in a structured manner and think about the connections between information. It corresponds to the Step 2 in Section 3.

DP2, **Contextual Reflection and Judgment**, obviously contains two parts: reflection and judgment. As a result, we categorize DP2 into two sub-codes:

- *Context Recall*: This sub-code indicates that we inspire the participants to connect their behaviors and preferences to their life habits and circumstances. It corresponds to the second half of Q1 and Q2, as well as Q3 and Q5 in Section 1; Q2 and Q5 in Section 2; and Q1, Q2, Q4, and Q5 in Section 3.
- *Evaluation and Judgment*: This sub-code means that we encourage the participants to express their demands for the YouTube recommendation system by evaluating the videos they watched and their behaviors. It corresponds to Q4 and Q6 in Section 1; Q3 and Q6 in Section 2; and Q3 and Q6 in Section 3.

DP3, **Guidance-Based Reflection**, serving as a supporting principle to DP1 (Fully Reflect by Slowing Down the Thinking Process) and DP2 (Contextual Reflection and Judgment), is present throughout the entire process. Unlike the other two design principles that have exact moments and stages where they can be clearly identified, DP3 (Guidance-Based Reflection) does not have specific and obvious instances. It is intertwined with the overall process and subtly guides and influences users' reflection, making it difficult to isolate or code.

Once the initial framework was established based on the criteria we mentioned above, two people were required to code the first 5 of the 20 groups independently without any discussion. Then we used the codes for intercoder reliability analysis (ICR), which was a measurement of how much coders agree when coding the same data [55]. For these 5 groups, we reached an agreement of 33.1% on values. In the following, we conducted a thorough discussion and aligned the codes one by one. We focused on what values participants' responses should correspond to, how to merge similar contextual factors, and which of the six categories each contextual factor falls into, thus refining the framework and ensuring a consistent understanding of the codebooks. When the agreement between two coders reached more than 98%, which meant we achieved a high level of reliability, we proceeded to manually code the remaining 15 groups with the refined codebooks.

During the process of coding, we needed to select the whole paragraph and assign the corresponding codes. Figure 4.2 shows an example of coding. In this example, we assigned *trending topics* as the contextual factor for the reason that ”there was an attempt of assassination of former president of the United States Donald Trump” was a current topic at that moment, which drove the participants to watch this video. Additionally, the participant also expressed that the video did not provide the required information, which corresponds to the value of *Knowledge, Informativeness*, referring to users’ demand for the capacity of the recommendation system to recommend items of interest to users and help users acquire relevant information. This further derives the value of *Usefulness*, referring to the user’s demand for recommended items and services to be useful and helpful.

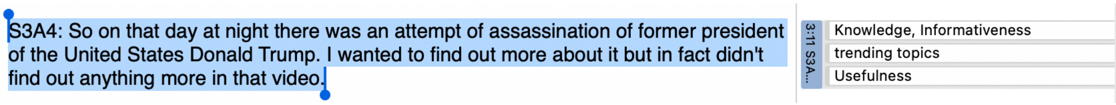

Figure 4.2: An Example of Coding

5 Results

This chapter aims to answer the second research question:

- *What contextual factors and personal values for participants' usage of YouTube are extracted during one's reflection?*

we begin by showing the participants' completion status of the experiment. The status at the time of submission and the corresponding number of participants are shown in Table 5.1.

Status at Submission	Number of Participants
Submitted with Completion Code (with video)	6
Submitted with Completion Code (no video)	14
Returned in Section 1	2
Returned in Section 3	4
Failed Attention Check	5

Table 5.1: Completion of the Experiment

As shown in the table, 20 participants submitted successfully with a correct completion code, with 6 of them managing to record the screen. 2 participants returned in Section 1, one of whom had minimal YouTube activity last week, resulting in very simplistic images generated by the data processing that required no self-reflection. Another participant, together with the 4 participants who returned in Section 3, was assumed to be unwilling to proceed with the study due to its complexity, time demands, or lack of interest. There were also 5 participants who failed the attention check. We set a popup window every two minutes on the first few pages to test whether the participants were still on our platform. If the participant missed the window twice, we assumed that they were not concentrated and failed the attention check.

During the process, participants were able to conduct the study simultaneously, proving the ability of our proposed approach to collect data on a large scale.

In the following, we will go into great detail to demonstrate the ability of the approach to ensure the depth of the collected data by discussing the results of the contextual factors and personal values that emerged from participants' narratives to reveal the significant themes that influence individuals' behaviors and user experiences and explore people's expectations for recommendation systems. In the following, we are going to investigate the co-occurrence between contextual factors and personal values to gain a deeper understanding of which values users prioritize in a specific context. Lastly, we will turn to the design principles to investigate their impact on the recalling and elicitation of contextual factors and personal values during the overall process.

5.1. Overall Analysis of Contexts and Values

5.1.1. Analysis of Contextual Factors

According to previous work [48], contextual factors can be divided into six dimensions, with which we can analyze the aspects that influence user interaction with the YouTube recommendation system. Each dimension can be further divided into several detailed contextual factors. Table 5.2 lists the description, quote example, and the number of each factor in all the documents.

Category	Contextual Factor	Description	Example Quote	Count
Application Context	Automatically played videos	YouTube's autoplay feature and random recommendations.	”It wasn't something I chose to watch. It was recommended by YouTube...”	9
	Video attributes	The specific characteristics and features of a video, including video title, author, content, and thumbnail.	”I watched the video because I was interested in the title and I wanted to find out what it was and widen my knowledge.”	40
Cognitive Context	Emotional needs	The state and conditions that should be fulfilled in order for people to experience happiness and peace.	”This helps me to unwind and re-boost my energy as well as calming down my overthinking capacity.”	28
	Personally interested topics	The content that users find appealing and valuable because of their hobbies, curiosity, jobs, related experiences, and educational backgrounds.	”I wanted to learn about developing my blog further, and I felt these videos about Squarespace were a starting point for beginners.”	76
	Repetitive watching	People want to watch the same videos over and over again.	”I wanted to laugh and comfort myself. I re-watched the video that I found hilarious the last time I watched.”	5
	Traits	Particular characteristics, qualities, and tendencies that someone has, such as personality and religious beliefs.	”All these are very good ways to strengthen one's faith.”	6
	Willing to connect to old times	People's willingness to reminisce the past times and experiences.	”I am satisfied with this video and music in the fact that it reminds me of my old high school days.”	2
Temporal Context	Stay outside	The physical location outside the house.	”I only listen to music when I go out (Tuesday and Friday)...”	6
	Watch on transport	The specific location of the transport vehicle, such as bus, train, or airplane.	”I listen to music the most when I travel by bus since it's time-consuming and boring.”	4
Social Context	Be alone	The state of being physically by oneself.	”...I always enjoy my lonely hours watching YouTube music videos.”	2
	Stay with others	Physically being in the company of other people.	”I enjoy listening to music and watching music videos, especially on a weekend when I am chilling with my friends.”	3
Task Context	Daily routines	One's life habits within one day.	”This was just a song to wake me up while having my morning coffee to get me in a good mood.”	29

	Do something while watching	When people engage with several tasks, such as watching videos while cooking, doing housework, and working, they will have a tendency for content that aligns with their ongoing tasks.	”I use it for entertainment and white noise so I can sleep or study or draw.”	25
	Job	An employment or a position in which someone works for a company or organization.	”I’m a developer, and I’m very interested in consumer electronics.”	4
	Personal planning	People’s long-term goal.	”Since my goals are to continue studying music and exploring that field to eventually do a Master’s or another Bachelor’s abroad, I guess that my watching habits are aligned to that.”	5
	Working Status	People’s job or working condition.	”Since I usually have less time (off work) to browse through YouTube, I just watch short clips and grasp everything I need in a jiffy.”	37
Temporal Context	Day of the week	The day type in one week, such as Friday, weekdays, and weekends.	”I find myself watching specific videos, like the Film and Animation videos, during the first half of the week because these are days I may be feeling like I need a sense of creative direction to help me plan out my week better, and the weekend time spent on the mentioned category is mostly for me to catch up on what I may have missed during the week.”	15
	Trending topics	The current news and popular fashions at a certain time.	”I mostly like to keep up with current affairs, so I watch a lot of news.”	12

Table 5.2: Overall result of contextual factors

Cognitive Context, which has 105 mentions, emerges as the most significant contextual factor. It refers to users’ cognitive abilities, attitudes [48], and internal mentality, which shape the way people perceive, engage with, and evaluate the content they encounter. Within this dimension, emotional needs and personally interested topics are mentioned the most frequently. They play a significant role in determining user preferences and behaviors, as they internally drive users to consume content. To be specific, people often look for content that can resonate with their current emotional state or help them reach the expected emotions. For example, users who feel stressed will probably seek content that can help them relax and forget their worries. Besides, people’s interests drive them to seek related types of content to pursue enjoyment, gain specific knowledge, and look for solutions, ultimately achieving personal goals. As such, the high frequency indicates that people turn to YouTube mostly for information and content they care about.

Task Context, with 88 mentions, is another key factor. It refers to people and objects surrounding the task, which offer advantages (“resources”) and drawbacks (“constraints”) for its achievement [48]. This context emphasizes how the tasks people perform simultaneously will influence their preferences and behaviors for the content they seek. The frequent mentions of this dimension show that many people tend to ”multitask”, such as their life habits in daily routines and their preferences for watching videos while doing other things. The nature of the ongoing task will dictate their choices for content, allowing them to complement their current status without requiring full attention, such as helping them focus on or be distracted from the task they are working on.

Application Context (49 mentions) refers to the representation of the application’s state and functioning [48], which means how the system presents content to users and the device the user is using to access the content. The high frequency of this context reflects that users are sensitive to how content is presented, such as video attributes, as this directly affects their visual impressions. People may be more inclined to click on videos of familiar authors or prefer videos with good editing and well-structured content.

Despite the fact that *Temporal* (27 mentions), *Physical* (10 mentions), and *Social Context* (5 mentions)occur less frequently, they are still relevant in several scenarios. Temporal Context means the time of day and date that is related to user activity patterns, such as the day of the week. People's preferences shifted between weekdays and weekends. There is a significant number of people who prefer to watch educational, short, or informative videos during the weekdays while turning to entertaining, causal, or long videos at weekends. Physical Context mainly focuses on users' physical environment at the time of content consumption. For instance, people may turn to more engaging videos in quiet places, while they watch some soft music videos in public places. Social Context refers to the influence of surrounding people, including relationship, dialogue, presence, and behavior [48]. People's preferences and behaviors can change greatly depending on whether they are alone or staying with others. Some people may feel comfortable exploring the content of their interests when being alone, while others enjoy sharing videos that interest them with friends.

User preferences and behaviors may be shaped and influenced by a combination of several factors. In addition, one type of contextual factors may influence another to some extent. The most prominent co-occurrence is between *Cognitive Context*, *Task Context*, and *Temporal Context*. This indicates that at different times, people typically have different ongoing tasks, and they will accordingly seek content that aligns with their current tasks as a way to fulfill their cognitive demands, such as emotional needs and personal interests. For example, during morning hours, people are probably working, and they are therefore more inclined to soothing videos to satisfy their emotional needs such as helping them calm down or stay more focused.

5.1.2. Analysis of Personal Values

In addition to contextual factors, we also focus on 5 themes of values relevant to recommendation systems, namely legal and human rights, public disclosure, safety, usefulness, and well-being. [65], each encompasses several values that users associate with their experiences on these platforms. Table 5.3 shows the description, quote example, and number of occurrences of all the values in each theme, providing insights into what users care about and prioritize when interacting with the YouTube recommendation system.

Theme	Value	Description	Example Quote	Count
Legal and Human Rights	Freedom of Expression	Platforms should not stop users from expressing their thoughts and opinions freely.	”I value the freedom and openness YouTube provides when it comes to self-expression.”	4
Public Disclosure	Accuracy (Factuality)	Platforms should make judgments and classify information correctly and prevent the spread of untrustworthy and unreliable information.	”I don't like deceiving people and I value honesty a great deal. In general, I dislike clickbait content...”	1
	Diversity	Platforms should satisfy people's desire to construct their views from various sources, which help them to make well-considered decisions [29].	”I felt less satisfied with...due to the lack of variety in content and repetitiveness of topics.”	9
	Knowledge, Informativeness	Platforms should satisfy people's demand for getting relevant information and being informed more effectively about the topics they care about [28].	”The video being educative should be the first and foremost of the quality every video should have.”	49
Safety	Safety, Security	Platforms should not contribute to increasing stress, anxiety, or a sense of being harassed by one's digital environment.	”People and blogs sometimes talk about things that can not align with my current situation. Sometimes they demotivate me.”	5
Usefulness	Agency, Autonomy, Efficacy	Platforms should enable users to act intentionally to achieve their goals [67].	”It (the watching habit) helps me...focus on what I am doing at a certain time.”	7
	Control	Platforms should allow users to control and customize their recommendations [59].	”I would love to watch the video that I chose rather than a random ad.”	6

	Efficiency	Platforms should allow users to achieve their goals efficiently.	”Since I usually have less time to browse through YouTube, I just watch short clips and grasp everything I need in a jiffy.”	8
	Usefulness	Platforms should provide useful service to users and help people find relevant items [35].	”I felt I wasted my time in watching it...It didn’t meet my expectations.”	34
Well-being	Accompany	Users’ demand for seeking content that provides a sense of companionship.	”I use it (YouTube) for entertainment and white noise so I can sleep, study or draw.”	16
	Community, Belonging	The feeling and faith that people want to be together and form a community through their commitment.	”It satisfies me because I now know that my fellow youth won’t just sit...”	2
	Connection	The individual’s need to establish and maintain meaningful relationships with others.	”Watching political news and debate help me inform my family and colleagues on the current state of world politics.”	9
	Entertainment	People’s demand for the capacity of recommendation systems to provide enjoyable and engaging content that adds pleasure to their free time.	”I use music as a gentle relaxation tool in times when I need to lie down and rest.”	40
	Inspiration, Awe	People actively seeking ideas, motivation, inspiration, and guidance that can help determine their choices and achievements [41].	”I watch...because these are days I may be feeling like I need a sense of creative direction to help me plan out my week better.”	19
	Mental Health	Platforms should discourage unhealthy types or amounts of use, and help people cope with stress in life.	”I learned that it’s best to keep a reasonable watch time.”	26
	Physical Health	Platforms should help users maintain and improve their physical health by providing information like healthy lifestyles.	”I think they are really accurate and it gives me some perspective on what habits or patterns I might have to change for myself.”	6
	Recognition, Acknowledg-ment	Platforms should enable people to recognize their worth and find identity in other people.	”I like learning about other people’s experiences in life.”	2
	Self-actualization, Personal Growth	Platforms should help them learn new things, reach their full potential, and realize a continuous improvement [41].	”I am quite satisfied because it didn’t say I watch more videos during my active working working hours...It means that I am putting my time into more important things for self-growth.”	37
	Self-identity	Platforms should enable users to understand and reinforce their identity through their interactions with them.	”It aligns with my personal goal because it helps me a great deal by...giving me a sense of belonging of who I am.”	4
	Well-being	Users should see content that can bring enjoyment both ephemerally and over the long term.	” I am satisfied with this video and music in the fact that it reminds me of my old high school days.”	20

Table 5.3: Overall result of personal values

Well-being was the most prominent theme, which has 137 mentions, underscoring its fundamental role in user experience. Many values, like *Mental Health*, *Self-actualization, Personal Growth*, and *Entertainment*, all strongly intersect with it. It is a complicated concept, as it refers to many objective aspects, like economic prosperity and employment rate, and subjective aspects, like mental health and fulfillment. In terms of recommendation systems, well-being refers to the ways in which individuals can be guided toward choices that enhance their overall quality of life. It has been explored as an important end goal for AI systems [14]. The high frequency of this theme indicates that users are increasingly seeking content that enriches their lives, promotes positive emotions, and fosters a sense of happiness. The growing emphasis on well-being reflects a shift towards intentionally pursuing meaningful interactions that align with their personal values and emotional needs in order to obtain creativity, learning, and connection.

Public Disclosure was the second most frequently mentioned theme, with 57 mentions. This theme reflects users' need for the quality, transparency, and diversity of the information they receive. People are increasingly aware of the importance of acquiring relevant, accurate, and comprehensive information that supports their decision-making and understanding across various domains [31]. Within this theme, *Knowledge, Informativeness* dominates other values, suggesting that users consider YouTube a crucial source of information. People increasingly turn to YouTube for knowledge and information across various domains, ranging from trending topics to specialized areas.

Usefulness (48 mentions) highlights users' demand for the practical benefits of the content they consume, and they turn to YouTube for the tangible utility of information and services. Recommendation systems should be useful to users, such as recommending appropriate items without a specific query. *Usefulness*, the most frequently mentioned value within this theme, suggests that users frequently seek content that can address specific issues and needs, and they prioritize whether YouTube can provide or recommend videos that can directly benefit their daily lives. *Control* and *Agency, Autonomy, Efficacy* are another two values that are closely related to the theme of usefulness. They represent users' demand for recommendation systems to allow flexible and free control of their preferences and enable them to choose the recommended content and find what they want through specific methods.

Safety (5 mentions) remains an essential concern despite the fact that it was much less frequently mentioned. This theme is important for maintaining trust in YouTube, as people expect the YouTube recommendation system to filter inappropriate content that could pose threats to their mental health and emotional well-being and ensure a reliable viewing environment. This is especially necessary for children, vulnerable populations, and groups with mental health problems.

The theme of *Legal and Human Rights*, with only 4 mentions, is rare in the context of the YouTube recommendation system, indicating that these concerns may not always be top of mind for users during casual content consumption, but it is still an aspect that needs to be taken into consideration. The values in this theme can be considered rights, which might not be frequently articulated because they are background concerns that are inherently expected. People may not be aware of these values during interactions because they tend to focus more on tangible experiences.

5.2. Co-occurred Values in Specific Contexts

The co-occurrence analysis between values can reveal how different values interact with each other on the YouTube recommendation system to affect people's preferences. Furthermore, understanding co-occurred values in specific contextual factors contributes to deeper insights into user preferences and behaviors. According to these patterns, we can investigate how different contexts exert an influence on people's prioritization of values for YouTube, thus helping YouTube provide more useful and personalized recommendations.

5.2.1. Co-occurrence of Entertainment and Knowledge, Informativeness

The two most frequently mentioned values, *Entertainment* (40 mentions) and *Knowledge, Informativeness* (49 mentions), are intertwined with each other. Their co-occurrence reveals an interesting relationship between these two seemingly contradictory values. While *Entertainment* is always associated with leisure and amusement, *Knowledge, Informativeness* is typically connected to learning and information. The intersection of the two values indicates that people have a preference for content that can both satisfy their need for enjoyment and enable them to be informed about information. This reflects people's desire for videos that can strike a balance between the educational element and the entertaining element. Videos should not only impart knowledge but also package themselves into an engaging and entertaining form. For example, P2 mentioned that:

> ”I feel satisfied (with the video) because I know I have gained more knowledge and the funny aspect of the video made it not so boring.” (P2)

Moreover, the occurrence of these two values is influenced by two recurring contexts. In the *Temporal Context* (27 mentions), the balance between *Entertainment* and *Knowledge, Informativeness* will shift depending on the day of the week. On weekdays, people may be preoccupied with their work, they therefore prioritize knowledge-focused videos and seek content that is educational or relevant to their professions and jobs. As such, the informative value might dominate. Whereas on weekends, people shift their focus to more entertaining content, helping them unwind after a busy week and create a more enjoyable experience.

> ”During the workweek, people often seek educational or professional content to make the most of their limited free time. On weekends, there is more free time, which encourages watching entertainment content.” (P14)

In the *Task Context* (88 mentions), the co-occurrence of these two values will vary based on the nature of the ongoing tasks. When people are engaged with activities that require concentration, like learning new technologies or preparing for exams, some might prioritize informative content that is highly relevant to their domain, or is straight to the point to get them informed about information effectively, while others prefer delightful content to relieve stress. However, when people are performing more routine tasks or seeking relaxation, like cooking or doing housework, someone may put more focus on entertainment and amusement, allowing them to enjoy the process or stay focused. Others, on the other hand, may prefer content that combines the two values to make the tasks more enjoyable and enable them to absorb information, thus getting more energetic and motivated.

> ”Also (I enjoy listening to music) while working as it relieves some stress at times.”
> ”I do it (play the music videos) basically for enjoyment and relaxation and sometimes just to dance a bit.” (P18)

5.2.2. Co-occurrence of Self-actualization, Personal Growth, Knowledge, Informativeness and Inspiration, Awe

The high frequency of mentions proves that *Self-actualization, Personal Growth* (37 mentions), *Knowledge, Informativeness* (49 mentions) and *Inspiration, Awe* (19 mentions) always intersect with each other when individuals attempt to seek personal fulfillment and development. People who are eager to realize their personal growth tend to prioritize their access to relevant information and proactively seek guidance and ideas.

> ”I’m satisfied because mostly it relates to learning things for the benefit of myself...I wanted to learn about developing my blog further, and I felt these videos about Squarespace were a starting point for beginners.” (P1)

The close relationship between those values largely depends on *Cognitive Context*, especially in terms of the personally interested topics, as it has 76 mentions. Individuals who are in pursuit of personal growth will probably turn to content that is related to their domains of interest to acquire inspiration and care about whether they can obtain information from the videos. In turn, people will seek content that aligns with their interests to get informed about the information they care about and actively acquire inspiration and ideas, thus using it as a tool to achieve personal development.

> ”I’m a creative and I spend most of my time doing research and learning new things on all things creative, especially digital creative related...I’m satisfied with the amount of information provided as this was the goal and fed my curiosity and hunger for anything related to the digital creative industry.” (P8)

5.2.3. Co-occurrence of Self-actualization, Personal Growth, Knowledge, Informativeness and Usefulness

The co-occurrence of *Self-actualization, Personal Growth* (37 mentions), *Knowledge, Informativeness* (49 mentions), and *Usefulness* (34 mentions) highlights the interplay between individual growth and the acquisition of relevant and useful information. Self-growth reflects an individual’s pursuit of developing to their full potential, driving them to seek content that empowers them to get informed about knowledge that can enhance their worldview. At the same time, people will prioritize content that has utility and practical applications in their personal development to help them solve problems, improve skills, and make informed decisions. On the other hand, people’s seeking for self-actualization is typically driven by their desire to expand their knowledge and obtain content with practical value. This implies that people have a preference for content that is both informative and useful to provide them with insights into solving various problems, which act as fuel to create a continuous cycle of self-improvement.

> ”I hate to watch the same thing over and over and I am more interested when I have something to learn from the video and it’s something interesting, like video essays and more informative videos that I never watched and that can manage to get my attention.” (P6)

These values usually occur together within *Task Context*, especially influenced by video attributes. These values are related to specific goals, such as learning new skills and obtaining knowledge. In this context, video attributes like title, author, content, and pace, play an essential role in attracting users. For example, people tend to prefer videos with eye-catching covers and titles that are relevant to their demands. Besides, video attributes work as an indicator, which influences users' satisfaction levels. When learning and working, individuals will pay attention to whether the video attributes can satisfy their needs. They probably prefer content that both aligns with their intrinsic motivation and provides insights relevant to their current tasks and goals.

> ”The purpose of watching it was to learn about possible new technologies. It felt like more of a marketing video than an interesting technology one. I didn't like the spin the video creator puts on the whole AI thing.” (P12)

5.2.4. Co-occurrence of Accompany and Mental Health

The co-occurrence of *Accompany* and *Mental Health* is also supported by quantitative data. From Table 5.3, *Accompany* appeared 16 times, and *Mental Health* appeared 26 times, indicating that both values are frequently discussed by participants. The frequency highlights the importance of the values in the YouTube recommendation system and shows a potential overlap between the two. In some cases, the two values do not inherently conflict but can compensate for each other. For instance, *Accompany* means providing a sense of companionship, with which people can cope with stress in life and pass their free time to relief from loneliness. This can somewhat support and maintain mental health.

> ”Whenever I am feeling lonely or thinking of something, I play music...This watching habit is cool to me because it helps me maintain my mental health and gives me energy and focus on what I am doing at a certain time.” (P4)
> ”When I watch videos, I want to hear another person speaking. It makes me feel that somebody is talking to me...I am pretty much satisfied with my watch habit because it helps me survive dark times and hardships. I find watching some content creators helpful when it comes to problems in my life. It makes me less sad and more stable.” (P9)

However, user experiences are complicated and sometimes reveal a contradiction between the values. To be specific, some users demand that the system fulfill the value of companionship to help them enjoy their leisure time, yet they simultaneously consider that long screen time conflicts with their pursuit of mental health. They enjoy the company brought by watching videos but also recognize that it can be time-consuming and goes against their values around maintaining their mental well-being. The co-occurrence of the two values indicates that people are facing a challenge in striking a balance between emotional needs and maintaining mental health.

> ”The blogs help me to pass the free time and are a really good source of company when I am eating and doing other stuff...I am not that satisfied with my patterns because it means that I am watching more videos when I should sleep, and I have been watching blogs too much, which means that I have been bored lately and to cope with that I spend more free time on YouTube.” (P6)

5.3. Analysis of Design Principles

Based on the guidelines for the division of design principles, DP1 (Fully Reflect by Slowing Down the Thinking Process) and DP2 (Contextual Reflection and Judgment) are central to this study. In this section, we are going to present the results of our investigation into the impact of the design principles applied to the task. The analysis is divided into two parts: quantitative analysis and qualitative analysis. By combining the two aspects, we can gain a comprehensive understanding of how these design principles contribute to users' reflections on their preferences and behaviors to generate insights into the contextual factors and values of YouTube.

5.3.1. Quantitative Analysis

The quantitative analysis demonstrates the numerical data measuring the role of each sub-code in enabling people to self-reflect. Figure 5.1 reveals the number of contextual factors and personal values motivated by each sub-code in the design principles.

		Contextual Factors 214	Legal and Human Rights 4	Public Disclosure 57	Safety 5	Usefulness 48	Well-being 137
Different Angles	27	18	0	1	0	1	4
Speculation and Comparison	130	59	0	8	0	2	14
Context Recall	194	144	0	31	2	23	62
Evaluation and Judgement	106	51	4	25	3	24	71

Figure 5.1: Numerical Data on the Role of the Design Principles

From this chart, it is clear that the sub-code *Context Recall* triggered 144 contextual factors, indicating that DP2 (Contextual Reflection and Judgment) could truly enable people to make connections to their experiences, habits, and circumstances, thus reflecting on the reasons behind their behaviors and preferences for watching YouTube videos. Besides, the sub-code *Evaluation and Judgment* elicits 127 values in total from 5 themes, showing that participants successfully reflect on their values and demands for YouTube with the guidance of our proposed design principles.

Moreover, the sub-code *Context Recall* was able to trigger about the same number of contextual factors and personal values, which we did not anticipate. It makes sense as these two are closely related. When people relate their watching patterns to the contexts, in addition to thinking about the contextual motivations that shaped their preferences and behaviors, they are equally reflecting on their expectations for YouTube and how these contexts align with their personal values. As such, *Context Recall* is able to facilitate dual-layered reflection, which highlights the importance of this sub-code in assisting people with recalling the external factors and conducting internal evaluations, making it a key sub-code in understanding how people engage with self-reflection.

5.3.2. Qualitative Analysis

The qualitative analysis delves deeper into how these design principles support and interact with each other, especially how DP1 (Fully Reflect by Slowing Down the Thinking Process) facilitates the process of *Context Recall*, how *Context Recall* boosts *Evaluation and Judgment*. For each part, we can break it down into several parts based on the corresponding sub-codes.

***Speculation and Comparison* facilitates the process of *Context Recall*.** The process of *Speculation and Comparison* plays a significant role in helping participants reflect. They began by making simple guesses about their favorite video categories and watching time periods. Once they started recalling, they naturally thought about the reasons behind their guesses by making connections to their interests, habits, and recent activities.

> ”(I prefer to watch videos related to) classical music, history and culture, and web design. I am currently building a web blog for myself and others that I intend to monetize. For history, culture, and politics, I like to listen to Johnny Harris and I like to listen to other stuff about languages, cities, and their histories. For classical music, I’m a classical musician.” (P1)
> ”I prefer to watch between 14:00 and 23:00. I am a night owl and I can get most of the work done at night when it is calm, and I am not a huge fan of brightness and loud noises which can be a case in the day.” (P10)

In the following, participants were provided with visualizations generated from their actual data. When they encountered similarities and differences between their expectations and the real situation, they would definitely carefully consider and interpret the reasons behind them. If their guesses aligned with the actual data, they would probably attribute it to their hobbies, current status, and life habits. On the other hand, those cases that were inconsistent with their speculation could easily arouse people’s interest and attention, leading them to wonder if some recent special events or unconscious behaviors were to blame so that they could have a deeper understanding of themselves.

> ”Almost accurate as well. This is the time I have lunch and the time after I have knocked off and gone home.” (P8)
> ”My guess does not really align with the actual data...So I learned about myself that maybe sometimes I am getting a little bit distracted in my work and focus myself on different kinds of things unrelated to the work I perform.” (P3)

***Different Angles* facilitates the process of *Context Recall*.** We provide two perspectives corresponding to Section 1, video categories, and time periods, allowing the participants to choose a more interesting perspective to continue exploring in depth. In most cases, people chose the angle they guessed better in Section 1. To be specific, participants tend to choose a perspective that they guessed more accurately before in the subsequent task, as they may feel more comfortable and confident when recalling the details. By switching between the two angles, they were able to find a direction they could connect to the previous speculation, thus conducting further explorations and reflections on contexts.

> ”It aligns perfectly; as I said I wake up at around 11:00 to 12:00 and watch more around 2:00 with slight peaks at 17:00 to 20:00...the blog helps me to pass the free time and is a really good source of company when I am eating and doing other stuff, and I watch video game stuff because I am generally interested more on that subject and watching more about games lately.” (P6)

Besides, there were some participants who selected the angle that they did not predict correctly in Section 1. This is mostly due to the fact that they considered the differences interesting and tried to find out the reasons. To be specific, a participant who made the wrong speculation may choose that angle in Section 2 to investigate what led to the incorrect assumption. The exploration into unexpected contexts facilitates a deeper and more detailed recall, as the participants were motivated to consider their cognitive process behind the speculation.

> ”It is totally different and could be due to the reason that I was only referring to the recent past and not the whole week...I can see that I have a lot of interest in listening to music right after I wake up and I can be focused on learning in the mornings. I think this can be due to the reason that I have more energy and focus in the hours after waking up and maybe not much interest in doing productive things once times start passing.” (P10)

Constructing a Story Map* facilitates the process of *Context Recall. The sub-code of *Constructing a Story Map* involves requiring participants to reconstruct their day by filling in the activities in the blocks according to the timeline. This process enables participants to connect their watching behaviors to their activities on that day. Based on the Miro Boards we collected, most participants managed to complete the timeline, indicating that our design can effectively and successfully guide people to recall through the day and facilitate detailed memory. Figure 5.2 shows an example of the constructed timeline from a participant.

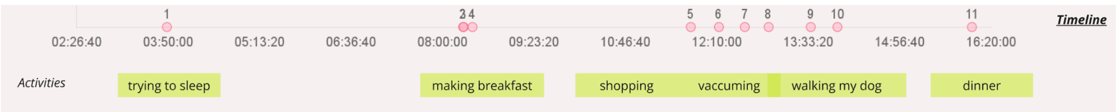

Figure 5.2: An Example of the Constructed Timeline

In addition to recalling the activities, participants were asked to choose 5 videos according to the reference table we provided and evaluate their satisfaction levels based on their viewing goals. They needed to fill in the watching purposes in the Miro Board, and then answer the related questions. Interestingly, in some cases, the activities and purposes they recalled in the Miro Board were consistent with the answers they provided in the subsequent reflection questions. The consistency reveals that the act of *Constructing a Story Map* can successfully enable participants to reflect on the reasons for their watching behaviors by linking to their activities and viewing goals. For instance, Figure 5.3 shows that P18 mentioned that he was busy with work from about 7:00 a.m., and he listened to some music. Correspondingly, he noted almost the same thing in his answers to the reflective questions:

> ”This was just a song to wake me up while having my morning coffee to get me in a good mood. The beat of the music video is extremely good and gives a nice upbeat rhythm to get you in the right spot to start a good day.” (P18)

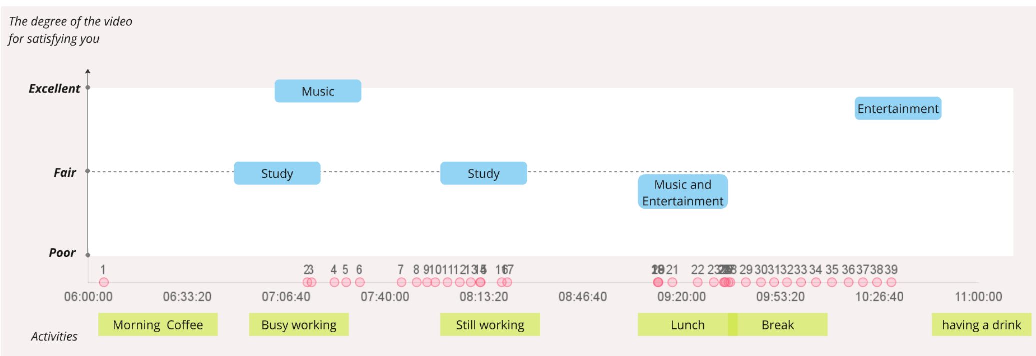

Figure 5.3: An Example of Participants’ Consistency in Constructing the Story Map and Answering Questions

Furthermore, we provide detailed information on the videos they watched, including title, author, watching time, and category, in the reference table. This helped participants recall the detailed video content, therefore reflecting on the reasons for their preferences.

> ”The scene of the video and how they were acting up their video as well as how it aligns with their lyrics. It is actually what I have been thinking when I was still young and in school.” (P4)
> ”I like the scene where the actor is singing and Bob Dylan walks up onto the sidewalk in the city in the 1960s probably. It is entertaining.” (P15)

Context Recall **boosts** ***Evaluation and Judgment***. By allowing participants to effectively recall the contexts, they are able to evaluate whether their watching behaviors and the recommended videos are consistent with their personal goals and expectations for YouTube. The process of *Context Recall* helps participants to judge whether their demands and pursuits for YouTube were satisfied or whether they were deviant, therefore reflecting more critically on their usage of YouTube.

> ”The top category is pretty spot on, as I am a classical musician and I tend to watch classical music mainly all the time...Since my goals are to continue studying music and exploring that field to eventually do a Master’s or another Bachelor’s abroad, I guess that my watching habits are aligned with that.” (P1)
> ”I mostly like to keep up with current affairs so I watch a lot of news. I also try to watch videos on social experiments and see how different people react to different situations. All of this helps me change my mind in a way that is beneficial to me while tackling real-world problems and different people I meet out there. I am satisfied with this because it shows I am willing to put in time to learn more about people and their behaviors. This helps me identify people who are worth my time and those who are willing to work with me to build a better future that benefits us both.” (P20)

6 Discussion

In this section, we are going to consider the implications of the results mentioned above and discuss their relevance to future practices. We first discuss the guidance for the YouTube recommendation system based on the results and analysis of contextual factors and values and then explore the lessons learned from our design to provide direction for the design of a future crowdsourcing approach with personal reflection integration.

6.1. Guidance for YouTube Recommendation System

The co-occurrence analysis of contextual factors and values reveals intricate relationships between different aspects that shape user preferences and behaviors. As such, understanding these relationships provides valuable insights and guidance for improving the YouTube recommendation system to better satisfy users' needs and expectations.

The high frequency of the co-occurrence between the *Cognitive Context* and *Temporal Context* indicates that time has a significant influence on users' cognitive states. Based on this pattern, the YouTube recommendation system can be enhanced by implementing strategies of dynamic personalization, which means the system can dynamically adjust the weight of cognitive states at different times. By exploring the temporal nuances of the user's cognitive states, the recommendation system can predict more relevant content to enhance the viewing experience. For example, during productivity hours, the weight of personally interested topics could be increased to recommend videos related to users' jobs and educational backgrounds, such as tutorials and industry insights. Conversely, for the time off work, the weight of emotional needs could be raised to recommend videos that help them relax and mitigate fatigue, such as soft music and comedies.

The strong co-occurrence between the *Cognitive Context* and the *Task Context* suggests that users' ongoing task will affect their cognitive states, thus turning them to content that can satisfy their emotional needs and personal interests. On the basis of this pattern, the recommendation system should be adaptive to users' current objectives in order to satisfy their cognitive demands. For example, if the system detects that users are executing tasks that require full attention, such as studying and working, it is supposed to recommend relevant content to align with their cognitive demands, such as satisfying people's emotional needs by recommending soothing videos that help them stay focused and feel relaxed, or fulfilling people's individual interests by recommending informative videos that help them get knowledge related to their job and study.

Based on the overall result of the values mentioned above, the theme of *Well-being* dominates all the themes. The emphasis on well-being reflects that YouTube should prioritize content that aligns with users' well-being goals, thereby encouraging engagement and leading to richer and more fulfilling experiences. Due to the fact that *Well-being* encompasses several dimensions, YouTube therefore needs to focus on different aspects, including facilitating deeper connections among users, promoting mental health, and providing resources for users' personal development. Additionally, the themes of *Public Disclosure* and *Usefulness* are also frequently mentioned, indicating that YouTube should consider enhancing its role as a platform for sharing relevant, accurate, diverse, and useful information. To be specific, YouTube should empower users to quickly and easily find reliable videos that provide practical value. Moreover, by promoting content that is credible, diverse, and rich, YouTube can strengthen its role as a trusted source that provides the required knowledge and helps people make informed decisions.

In addition to independent analysis of these themes, the co-occurrence of values under specific contexts is worth exploring to provide insights that can guide the enhancement of the YouTube recommendation system. The co-occurrence of *Entertainment* and *Knowledge, Informativeness* is an interesting phenomenon, as these two values are somewhat contradictory. People see YouTube as a tool for relaxation while also wanting to gain knowledge from it. This highlights users' preference for content that both satisfies their need for enjoyment and enables them to be informed about information. Based on this trend, YouTube should blur the line between pure entertainment and informative demand, and recommend educative videos packaged in an entertaining form, such as gamified learning and edutainment. This can meet users' needs for relaxation and obtaining knowledge at the same time. Users can maximize the value of their leisure time while consuming content for entertainment and relaxation. Furthermore, people's preferences for *Entertainment* and *Knowledge, Informativeness* shift depending on the *Temporal Context* and Task Context. Therefore, the YouTube recommendation system can incorporate time patterns into its algorithm to recommend corresponding categories of videos on different days or time periods. Additionally, YouTube can refine recommendations by recognizing users' current tasks. This could be done by adding a real-time feedback loop. By prompting users to indicate their ongoing tasks, YouTube can adjust the recommendations based on users' immediate needs.

The frequent co-occurrence of *Self-actualization, Personal Growth*, *Knowledge, Informativeness*, and *Usefulness* indicates that users who are eager to achieve their goals and realize personal growth may prioritize their access to relevant information and the usefulness of the obtained knowledge. They usually seek content that positively and directly contributes to their personal development. As such, YouTube could identify users' growth-oriented patterns through their watching history and previous interactions and recommend relevant videos. For example, YouTube should recommend videos of specialized knowledge and skills to users who have a personal plan to further their studies. Moreover, it could suggest and adjust the content tailored to users according to their feedback in order to make sure the recommended content is useful both ephemerally and over the long term. In addition, these values are often affected by the *Task Context*, especially the video attributes, as video title, author, and content play a significant role in attracting people who are pursuing personal growth. The current YouTube interface only has the title and author of the video, and when hovering over the video cover, the video will play automatically. However, it is inefficient, since people cannot get the key information and judge whether the video is useful or not in a short period of time. As such, YouTube should provide transparent and meaningful descriptions of the video when people hover over the cover, allowing them to get a quick and brief overview and determine whether to watch or not.

The co-occurrence of *Accompany* and *Mental Health* suggests that when people are watching videos that provide a sense of presence, they tend to feel comforted since it can help them cope with life stress or pass their free time. As a result, YouTube can integrate the real-time feedback mechanism into its algorithm to recognize when users feel lonely or stressed and then recommend corresponding videos. However, despite the fact that a sense of company contributes to people's mental well-being, the findings also suggest that excessive screen time due to a desire for companionship may conflict with users' mental health goals. To solve this conflict, YouTube can add a gentle notification system. When users' screen time exceeds certain limits, the system can send an alarm or provide recommendations for healthier content choices, thus helping people strike a balance between *Accompany* and *Mental Health*.

6.2. Directions for Future Design

The quantitative and qualitative results demonstrate the efficacy of design principles in encouraging people to make personal reflections on their preferences and behaviors, with some specific sub-codes playing a crucial role in guiding participants to consider the contextual factors and elicit personal values.

The data indicates that the sub-code *Context Recall* is the key element of motivating people to self-reflect. Due to the fact that this sub-code triggered an almost equal number of contextual factors and personal values, future design should focus on enhancing this dual-layered reflection. One possible solution is to reinforce users' understanding of the connection between what shapes their behavior and preferences and their personal values when they are recalling contexts.

The qualitative result also underscores the importance of DP1 (Fully Reflect by Slowing Down the Thinking Process) in assisting people with recalling contextual factors. This fact indicates that future applications should take mechanisms that encourage thoughtful reflection into consideration. Some predefined and designed points should be included in order to prompt users to slow down their thinking process and reflect seriously on themselves.

More specifically, *Speculation and Comparison* together with visualizations enable participants to compare their expectations with actual data in a straightforward way. This process takes time, and therefore participants are able to pause for serious reflection. Besides, visual aids also play an important role in the reflection process, especially when people need to make comparisons, as visualizations can be more intuitive than data. As such, the process of speculation and comparison and visual cues can be applied to make reflections on the usage of other applications. For instance, in Music apps such as Spotify, people can guess their patterns of listening to music and then see a clear comparison between their expectations and their actual performance generated from the historical data, thus gaining a better understanding of themselves.

Providing participants with *Different Angles* can encourage more detailed recall and deeper reflection. The ability to switch between different perspectives enables people to select an angle with special patterns or that they are more interested in. Based on the success of this sub-code, it can be incorporated into the reflection study on the usage of other applications as well, allowing people to investigate their behaviors from multiple perspectives. In addition to containing some external angles like time and categories, some internal factors, such as emotions and pursuits, can also be taken into account. Empowering people with the freedom to select their way of reflection will encourage more meaningful and thoughtful insights.

The process of *Constructing a Story Map* can significantly facilitate people to recall carefully and deeply. By reconstructing their day step by step according to guidance, people can slowly connect their watching behaviors to their activities and events, allowing them to reflect in an interesting and interactive way. Those reflection tools enable people to recall their past experiences in a structured manner and identify significant moments. As a result, future work can consider integrating those tools into studies that require people to recall past behaviors. For instance, people can track their journeys or capture the relationships between their emotions and actions by constructing timelines or maps.

6.3. Limitations

Our proposed approach addresses the inability of current studies to simultaneously address the breadth and depth of user feedback collected. However, there is still a long way to go. In this section, we highlight some areas for improvement and suggest directions for future work.

One limitation of our approach is that it merely focused on analyzing users' textual answers and ignored other insightful actions, such as users' voice and their interactions with the screens. To be specific, some people will "think aloud" during the process, i.e., they will express their thoughts to themselves during the reflection process. On the other hand, how people interact with the interface, such as scrolling, pausing, and switching between different angles, can also provide insights into their reflection patterns. At first, I tried to take these aspects into consideration and included a screen recording function to capture the entire interaction during the reflection process. Nevertheless, this function had some problems. Technical issues such as users refreshing the web pages and losing network connectivity will interrupt recording, leading to the failure of the screen recordings to be successfully uploaded to the server. As a result, it was difficult for participants to complete screen recordings, leading us to collect only a small number of videos. Besides, we also found it complicated to analyze the video content, so our analysis covered only the textual part.

Based on this limitation, future work could combine different forms of people's behaviors for analysis to gain a more comprehensive understanding of their reflection process. This may include integrating audio recordings to analyze participants' verbalization, and screen recordings to investigate how they complete the task.

According to the fact that some participants stopped partway through the study, another limitation of our approach is related to fatigue. The structure of our design asks participants to make personal reflections and provide large amounts of written responses, which requires concentrated attention and detailed consideration. This will result in people thinking it is boring and overloaded. Additionally, our design of the task may limit participants' freedom and creativity to reflect to some extent. To be specific, we provide the participants with predefined frameworks for the reflection process, which may restrict their natural thinking process. Some people may prefer more flexible and interactive tasks, allowing them to reflect and express themselves in a personalized and comfortable way, thus enhancing engagement and encouraging a deeper and more natural self-reflection.

To address the limitation, future work can focus on increasing task flexibility to reduce participant fatigue and facilitate engagement. One potential is the introduction of artificial intelligence agents. They can provide real-time feedback based on the progress and engagement of participants and act as a companion throughout the entire process. For example, they can adjust the difficulty or style of the question according to participants' status and guide participants to conduct reflection through attractive and interactive discussions. Additionally, incorporating gamification elements into reflection may make the process more appealing and engaging as well.

7 Conclusion

This research is motivated by the fact that recommendation systems sometimes cannot make correct predictions of users' preferences, resulting in a degraded user experience. This is due to recommendation systems not acquiring sufficient user feedback to improve the quality of recommendations. To be specific, most mechanisms of collecting user feedback are unable to delve into deeper reasons behind people's preferences and behaviors. Instead, they merely focus on numerical scales or binary indicators, such as ”like” and ”dislike”, which are too simple and will restrict people's expressions.

Actually, there are many exact reasons behind people's behaviors of clicking on the ”like” and ”dislike” buttons, such as interests, habits, and current status, which are called contextual factors. Besides, there are other aspects affecting user experience. In our research, we only focus on personal values, referring to what people consider important in recommendation systems to assist them with achieving their pursuits and goals. Many methods have been proposed to collect contextual factors and personal values, such as offline interviews and crowdsourcing tasks. However, most of them have limitations, indicating that few studies now can satisfy both the depth and breadth of the data collected. As such, there is a research gap in how to collect deep user feedback on a large scale.

In order to bridge the gap, we introduce personal reflection. It refers to the process of taking time to think about and evaluate one's personal information, such as behaviors, cognitive patterns, attitudes, motivations, and past experiences. Many ideas have been presented, such as slow technology, gamification elements, and storytelling. Realizing the potential of personal reflection, we propose such a hypothesis, that integrating personal reflection into traditional crowdsourcing tasks can enable users to reflect on their personal data, thus generating more nuanced insights into their preferences and behaviors.

Based on this hypothesis, we present two research questions:

- *How to design personal reflection tasks in order to gather users' contextual factors and personal values for the YouTube recommendation system in the context of crowdsourcing?*
- *What contextual factors and personal values for participants' usage of YouTube are extracted during one's reflection?*

In order to address the first research question, we presented a novel crowdsourcing approach with personal reflection integration that can facilitate people to reflect on contextual factors and personal values of their usage of YouTube. The design of this approach follows several principles: allowing participants to fully self-reflect by slowing down their thinking process and to conduct contextual reflection and judgment under guidance.

To solve the second research question, we conducted a user study involving 20 participants recruited on the Prolific platform. We gained a qualitative and quantitative understanding of contextual factors, personal values, and the process of reflection based on design principles. The results confirmed the effectiveness of our approach, demonstrating its ability to collect deep user feedback, which is contextual factors and personal values, in a crowdsourcing paradigm. In addition, the result also reveals an analysis of contextual factors and personal values extracted from participants' textual answers, the co-occurred values in specific contexts, together with the role of design principles in the reflection process. These findings shed light on the implications of the insights generated from the result, providing guidance for the YouTube recommendation system and pointing to the directions for the design of similar studies in the future.

References

[1] Gregory D Abowd et al. "Towards a better understanding of context and context-awareness". In: *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1*. Springer. 1999, pp. 304–307.

[2] Anne Adams and Anna L Cox. *Questionnaires, in-depth interviews and focus groups*. Cambridge University Press, 2008.

[3] Juliana Alvarez et al. "An enriched customer journey map: how to construct and visualize a global portrait of both lived and perceived users' experiences?" In: *Designs* 4.3 (2020), p. 29.

[4] Xavier Amatriain et al. "Rate it again: increasing recommendation accuracy by user re-rating". In: *Proceedings of the third ACM conference on Recommender systems*. 2009, pp. 173–180.

[5] T Azungah. *Qualitative research: deductive and inductive approaches to data analysis. Qual Res J.(2018) 18: 383–400. doi: 10.1108*. Tech. rep. QRJ-D-18-00035.

[6] Eric PS Baumer et al. "Reviewing reflection: on the use of reflection in interactive system design". In: *Proceedings of the 2014 conference on Designing interactive systems*. 2014, pp. 93–102.

[7] Benjamin B Bederson. "PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps". In: *Proceedings of the 14th annual ACM symposium on User interface software and technology*. 2001, pp. 71–80.

[8] Rita Borgo et al. "Crowdsourcing for information visualization: Promises and pitfalls". In: *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments: Dagstuhl Seminar 15481, Dagstuhl Castle, Germany, November 22–27, 2015, Revised Contributions*. Springer. 2017, pp. 96–138.

[9] Alan Borning and Michael Muller. "Next steps for value sensitive design". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, pp. 1125–1134.

[10] Daren C Brabham. "Crowdsourcing as a model for problem solving: An introduction and cases". In: *Convergence* 14.1 (2008), pp. 75–90.

[11] Daragh Byrne and Gareth JF Jones. "Creating stories for reflection from multimodal lifelog content: an initial investigation". In: (2009).

[12] Eun Kyoung Choe et al. "Understanding quantified-selfers' practices in collecting and exploring personal data". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, pp. 1143–1152.

[13] Eun Kyoung Choe et al. "Understanding self-reflection: how people reflect on personal data through visual data exploration". In: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2017, pp. 173–182.

[14] IEEE Standards Committee et al. *IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being: IEEE Standard 7010-2020*. IEEE, 2020.

[15] Dan Cosley et al. "Is seeing believing? How recommender system interfaces affect users' opinions". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2003, pp. 585–592.

[16] Sandra Garcia Esparza, Michael P O'Mahony, and Barry Smyth. "Mining the real-time web: a novel approach to product recommendation". In: *Knowledge-Based Systems* 29 (2012), pp. 3–11.

[17] Stephen T Fife and Jacob D Gossner. "Deductive qualitative analysis: Evaluating, expanding, and refining theory". In: *International Journal of Qualitative Methods* 23 (2024), p. 16094069241244856.

[18] Rowanne Fleck and Geraldine Fitzpatrick. “Reflecting on reflection: framing a design landscape”. In: *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 2010, pp. 216–223.

[19] Susan G Forneris and Cynthia J Peden-McAlpine. “Contextual learning: A reflective learning intervention for nursing education”. In: *International journal of nursing education scholarship* 3.1 (2006).

[20] Batya Friedman, David G Hendry, Alan Borning, et al. “A survey of value sensitive design methods”. In: *Foundations and Trends® in Human–Computer Interaction* 11.2 (2017), pp. 63–125.

[21] Qin Gao and Stephan Vogel. “Consensus versus expertise: a case study of word alignment with mechanical turk”. In: *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon’s mechanical turk, CSLDAMT*. Vol. 10. 2010, pp. 30–34.

[22] Michael D Greenberg, Matthew W Easterday, and Elizabeth M Gerber. “Critiki: A scaffolded approach to gathering design feedback from paid crowdworkers”. In: *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 2015, pp. 235–244.

[23] Lars Hallnäs and Johan Redström. “Slow technology–designing for reflection”. In: *Personal and ubiquitous computing* 5 (2001), pp. 201–212.

[24] Larissa Hammon and Hajo Hippner. “Crowdsourcing”. In: *Business & Information systems engineering* 4 (2012), pp. 163–166.

[25] E Hassan et al. “Feedback Recommendation System Based on Structured Feedback Acquisition”. In: *Journal of Physics: Conference Series*. Vol. 1447. 1. IOP Publishing. 2020, p. 012051.

[26] Saskia Haug, Tim Rietz, and Alexander Maedche. “Accelerating deductive coding of qualitative data: An experimental study on the applicability of crowdsourcing”. In: *Proceedings of Mensch und Computer 2021*. 2021, pp. 432–443.

[27] Chen He, Denis Parra, and Katrien Verbert. “Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities”. In: *Expert Systems with Applications* 56 (2016), pp. 9–27.

[28] Natali Helberger. “On the democratic role of news recommenders”. In: *Algorithms, automation, and news*. Routledge, 2021, pp. 14–33.

[29] Natali Helberger, Kari Karppinen, and Lucia D’acunto. “Exposure diversity as a design principle for recommender systems”. In: *Information, communication & society* 21.2 (2018), pp. 191–207.

[30] María Hernández-Rubio, Iván Cantador, and Alejandro Bellogín. “A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews”. In: *User Modeling and User-Adapted Interaction* 29.2 (2019), pp. 381–441.

[31] Jennifer L Hochschild and Katherine Levine Einstein. “Do facts matter? Information and misinformation in American politics”. In: *Political Science Quarterly* 130.4 (2015), pp. 585–624.

[32] Jeffrey Hood, Elizabeth Sall, and Billy Charlton. “A GPS-based bicycle route choice model for San Francisco, California”. In: *Transportation letters* 3.1 (2011), pp. 63–75.

[33] Yifan Hu, Yehuda Koren, and Chris Volinsky. “Collaborative filtering for implicit feedback datasets”. In: *2008 Eighth IEEE international conference on data mining*. Ieee. 2008, pp. 263–272.

[34] Samuel Huron et al. “Constructive visualization”. In: *Proceedings of the 2014 conference on Designing interactive systems*. 2014, pp. 433–442.

[35] Dietmar Jannach and Gediminas Adomavicius. “Recommendations with a purpose”. In: *Proceedings of the 10th ACM conference on recommender systems*. 2016, pp. 7–10.

[36] Dietmar Jannach et al. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[37] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. “Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback”. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4.2 (2014), pp. 1–26.

[38] Mukund Jha et al. “Corpus creation for new genres: A crowdsourced approach to PP attachment”. In: *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon’s mechanical turk*. 2010, pp. 13–20.

[39] Yuan Jin et al. “Winning by learning? Effect of knowledge sharing in crowdsourcing contests”. In: *Information Systems Research* 32.3 (2021), pp. 836–859.

[40] V Kavinkumar et al. “A hybrid approach for recommendation system with added feedback component”. In: *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. 2015, pp. 745–752.

[41] Lianne Kerlin. “Human values: understanding psychological needs in a digital age”. In: *BBC R&D White Paper* 371 (2020).

[42] Nam Wook Kim et al. “Dataselfie: Empowering people to design personalized visuals to represent their data”. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–12.

[43] Bart P Knijnenburg et al. “Explaining the user experience of recommender systems”. In: *User modeling and user-adapted interaction* 22 (2012), pp. 441–504.

[44] Joseph A Konstan et al. “Grouplens: Applying collaborative filtering to usenet news”. In: *Communications of the ACM* 40.3 (1997), pp. 77–87.

[45] Miia Kosonen et al. “User motivation and knowledge sharing in idea crowdsourcing”. In: *International Journal of Innovation Management* 18.05 (2014), p. 1450031.

[46] Georgia Koutrika et al. “Courserank: A closed-community social system through the magnifying glass”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 3. 1. 2009, pp. 98–105.

[47] Albrecht Kurze et al. “Guess the data: data work to understand how people make sense of and use simple sensor data from homes”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing systems*. 2020, pp. 1–12.

[48] Carine Lallemand and Vincent Koenig. “Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS)”. In: *Proceedings of the 11th nordic conference on human-computer interaction: shaping experiences, shaping society*. 2020, pp. 1–13.

[49] Fritz Lekschas et al. “Ask Me or Tell Me? Enhancing the Effectiveness of Crowdsourced Design Feedback”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–12.

[50] Guoliang Li et al. “Crowdsourced data management: A survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 28.9 (2016), pp. 2296–2319.

[51] Ian Li, Anind Dey, and Jodi Forlizzi. “A stage-based model of personal informatics systems”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010, pp. 557–566.

[52] Nathan N Liu et al. “Unifying explicit and implicit feedback for collaborative filtering”. In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. 2010, pp. 1445–1448.

[53] Sean M McNee, John Riedl, and Joseph A Konstan. “Making recommendations better: an analytic model for human-recommender interaction”. In: *CHI’06 extended abstracts on Human factors in computing systems*. 2006, pp. 1103–1108.

[54] Elisa D Mekler, Ioanna Iacovides, and Julia Ayumi Bopp. “” A Game that Makes You Question...” Exploring the Role of Reflection for the Player Experience”. In: *Proceedings of the 2018 annual symposium on computer-human interaction in play*. 2018, pp. 315–327.

[55] Cliodhna O’Connor and Helene Joffe. “Intercoder reliability in qualitative research: debates and practical guidelines”. In: *International journal of qualitative methods* 19 (2020), p. 1609406919899220.

[56] Alina Pommeranz et al. “Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process”. In: *User Modeling and User-Adapted Interaction* 22 (2012), pp. 357–397.

[57] Alina Pommeranz et al. “Elicitation of situated values: need for tools to help stakeholders and designers to reflect and communicate”. In: *Ethics and Information Technology* 14.4 (2012), pp. 285–303.

[58] Zachary Pousman, John Stasko, and Michael Mateas. “Casual information visualization: Depictions of data in everyday life”. In: *IEEE transactions on visualization and computer graphics* 13.6 (2007), pp. 1145–1152.

[59] Pearl Pu, Li Chen, and Rong Hu. “A user-centric evaluation framework for recommender systems”. In: *Proceedings of the fifth ACM conference on Recommender systems*. 2011, pp. 157–164.

[60] Will Reese. “Nginx: the high-performance web server and reverse proxy”. In: *Linux Journal* 2008.173 (2008), p. 2.

[61] Ravi S Sharma, Aijaz A Shaikh, and Eldon Li. “Designing Recommendation or Suggestion Systems: looking to the future”. In: *Electronic Markets* 31 (2021), pp. 243–252.

[62] Junge Shen, Cheng Deng, and Xinbo Gao. “Attraction recommendation: Towards personalized tourism via collective intelligence”. In: *Neurocomputing* 173 (2016), pp. 789–798.

[63] Brigitte Smit. “Atlas. ti for qualitative data analysis”. In: *Perspectives in education* 20.3 (2002), pp. 65–75.

[64] Shahab Saquib Sohail, Jamshed Siddiqui, and Rashid Ali. “User feedback based evaluation of a product recommendation system using rank aggregation method”. In: *Advances in intelligent informatics*. Springer. 2015, pp. 349–358.

[65] Jonathan Stray et al. “Building human values into recommender systems: An interdisciplinary synthesis”. In: *ACM Transactions on Recommender Systems* 2.3 (2024), pp. 1–57.

[66] Alice Thudt et al. “Self-reflection and personal physicalization construction”. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–13.

[67] Lav R Varshney. “Respect for human autonomy in recommender systems”. In: *arXiv preprint arXiv:2009.02603* (2020).

[68] Fernanda B Viegas et al. “Manyeyes: a site for visualization at internet scale”. In: *IEEE transactions on visualization and computer graphics* 13.6 (2007), pp. 1121–1128.

[69] Maksims Volkovs and Guang Wei Yu. “Effective latent models for binary feedback in recommender systems”. In: *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 2015, pp. 313–322.

[70] Luis Von Ahn and Laura Dabbish. “Designing games with a purpose”. In: *Communications of the ACM* 51.8 (2008), pp. 58–67.

[71] Luis Von Ahn and Laura Dabbish. “Labeling images with a computer game”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 319–326.

[72] Luis Von Ahn, Mihir Kedia, and Manuel Blum. “Verbosity: a game for collecting common-sense facts”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, pp. 75–78.

[73] René TA Wietsma and Francesco Ricci. “Product Reviews in Mobile Decision Aid Systems.” In: *PERMID*. 2005, pp. 15–18.

[74] Zhaojun Yang et al. “Collection of user judgments on spoken dialog system with crowdsourcing”. In: *2010 IEEE Spoken Language Technology Workshop*. IEEE. 2010, pp. 277–282.

[75] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. “A survey of crowdsourcing systems”. In: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE. 2011, pp. 766–773.

[76] Ying Zhang, Xianghua Ding, and Ning Gu. “Understanding fatigue and its impact in crowdsourcing”. In: *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*. IEEE. 2018, pp. 57–62.

[77] Jian Zhao et al. "PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media". In: *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2014, pp. 203–212.

[78] Qian Zhao et al. "Explicit or implicit feedback? Engagement or satisfaction? A field experiment on machine-learning-based recommender systems". In: *Proceedings of the 33rd Annual ACM symposium on applied computing*. 2018, pp. 1331–1340.

[79] Qian Zhao et al. "Interpreting user inaction in recommender systems". In: *Proceedings of the 12th ACM Conference on Recommender Systems*. 2018, pp. 40–48.

[80] Feifei Zheng et al. "Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions". In: *Reviews of Geophysics* 56.4 (2018), pp. 698–740.

[81] Philip Zigoris and Yi Zhang. "Bayesian adaptive user profiling with explicit & implicit feedback". In: *Proceedings of the 15th ACM international conference on Information and knowledge management*. 2006, pp. 397–404.

[82] John Zimmerman et al. "Field trial of tiramisu: crowd-sourcing bus arrival times to spur co-design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, pp. 1677–1686.