To the Graduate Council:

I am submitting herewith a dissertation written by Senjuti Dutta entitled "Understanding Traits to Support Crowdworkers' Flexibility." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Dr. Scott Ruoti, Major Professor

We have read this dissertation
and recommend its acceptance:

Dr. Alex Williams

Dr. Anastasia Kuzminykh

Dr. Audris Mockus

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

To the Graduate Council:

I am submitting herewith a dissertation written by Senjuti Dutta entitled "Understanding Traits to Support Crowdworkers' Flexibility." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Dr. Scott Ruoti, Major Professor

We have read this dissertation
and recommend its acceptance:

Dr. Alex Williams

Dr. Anastasia Kuzminykh

Dr. Audris Mockus

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Understanding Traits to Support Crowdworkers' Flexibility

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Senjuti Dutta

August 2024

I would like to dedicate this thesis to my mother, Jyoti Dutta, whose infinite love and unwavering belief in my dreams have been my anchor and my guide.

Acknowledgments

I am immensely grateful to my advisor, Dr. Scott Ruoti, for his invaluable guidance and unwavering support throughout the course of this research. His expertise, insight, and constant encouragement have been fundamental to the success of this work. My sincere appreciation goes to the members of my dissertation committee: Dr. Anastasia Kuzminykh, Dr. Alex Williams, and Dr. Audris Mockus. Special thanks to Dr. Anastasia Kuzminykh and Dr. Alex Williams, who have also been my collaborators. Their collaborative efforts, constructive feedback, and insightful suggestions have been crucial in refining this research. I am also thankful to Dr. Rhema Linder, my collaborator, for his invaluable support and feedback to improve my research. His partnership has been essential in advancing my research to its fullest potential. I am thankful for the support provided by the National Science Foundation under award number CNS-2238001, which has been instrumental in the completion of this work. I owe a debt of gratitude to my labmates for creating an enriching and supportive environment in the lab. The collaboration, discussions, and shared moments have been a source of joy and inspiration throughout my PhD journey. To my friends in Knoxville, thank you for all the memorable moments and for providing a comforting diversion from the demands of my research. Finally, I express my deepest gratitude to my parents Jyoti and Swapan Dutta , my fiancée Anirban Roy, and my brother Sankhadip Dutta for their unwavering support and encouragement.

Abstract

Crowdworkers are drawn to the profession in part due to the flexibility it affords. However, the current design of crowdsourcing platforms limits this flexibility. Therefore, it is important to support the overall flexibility of crowdworkers. Incorporating a variety of device types in the workflow plays an important role in supporting the flexibility of crowdworkers, however each device type requires a tailored workflow. The standard workflow of crowdworkers consists of stages of work such as managing and completing tasks. I hypothesize that different devices will have unique traits for task completion and task management. Therefore in this dissertation, I explore what those traits are. Future work can build upon this research by creating tailored workflows and interfaces to best support each device type. To achieve this, this dissertation introduces four pivotal innovations : (1) understanding traits of task completion on smartphones to support the tailored workflow on smartphones in crowdwork (2) understanding of crowdworkers' current task completion and task management practices and expectations when working on smartphone, tablet, speaker and smartwatch to support the flexibility of crowdworkers on all these devices based on crowdworkers' work practices and expectations. (3) After a broad understanding of crowdworkers' practices and expectations across different devices, this thesis identifies the systematic differences among crowdworkers in order to develop customizable support depending on workers' individual need for flexibility in crowdsourcing platforms. (4) Finally, this dissertation looks into other popular crowdsourcing platform named Prolific to understand work practices of Prolific workers as well as compare Prolific with Amazon MTurk to gain a comprehensive understanding of the traits that support flexibility in different crowdsourcing environments.

Table of Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.1.1 Research Goal . . . 3

1.2 Research Contributions . . . 4

2 Background and Related Works **6**

2.1 The Desktop-Centric Nature of Crowdwork and Workstation's Relationship with Flexibility . . . 6

2.1.1 Non-Workstation Device Usage in HCI and Its Relationship with Flexibility . . . 9

2.2 Crowdwork on Non-Desktop Devices . . . 11

2.3 Relation to This Work . . . 12

3 Understanding the traits to support crowdwork completion on mobile devices **14**

3.1 Introduction . . . 14

3.2 Taxonomy Generation - Step 1: Targeted Literature Analysis . . . 17

3.2.1 Procedure . . . 17

3.2.2 Findings . . . 19

3.3 Taxonomy Support - Step 2: Mechanical Turk Support Survey . . . 23

3.3.1 Method: Online Survey . . . 23

3.3.2 Findings . . . 26

3.4 Taxonomy Demonstration - Step 3: An Analysis Of HITs In-The-Wild . . . 33

3.4.1 Method: Web Scraping and Taxonomy Application 33
3.4.2 Findings 36
3.5 Discussion 45
3.5.1 Implications for Design: A Taxonomy for Practicing "Good" Mobile Usability 46
3.5.2 Toward a Usable Practice of Cross-Device Crowdwork 47
3.5.3 The Frontier of Mobile Tooling in Crowdwork 49
3.6 Limitation 50

4 Understanding crowdworkers' perspective for using various device types in crowdwork 51
4.1 Introduction 51
4.2 Study Design 53
4.2.1 Survey Design 53
4.2.2 Study Execution 55
4.2.3 Analysis Methods 55
4.2.4 Participants 56
4.3 Results 58
4.3.1 Current HIT Completion Practices 58
4.3.2 Current HIT Management Practices 61
4.3.3 Desired Usage 65
4.4 Discussion 71
4.4.1 Recommendations for HIT Management 71
4.4.2 Recommendations for HIT Completion 73
4.5 Limitation 76

5 Understanding systematic differences between crowdworkers based on their work practices 77
5.1 Introduction 77
5.2 Methodology 79
5.2.1 Analysis Method 79

5.3 Result . . . 85
5.3.1 Description of each PCA Component . . . 85
5.3.2 Description of Clusters of Crowdworkers . . . 86
5.4 Discussion . . . 93
5.4.1 Design Guidelines based on Clusters of Crowdworkers . . . 93
5.4.2 Understanding Crowdworkers through the Lens of Traditional and Established Worker Groups . . . 96
5.5 Limitation . . . 97

6 Understanding platform-specific comparison of traits for supporting flexibility of crowdworkers 99
6.1 Introduction . . . 99
6.2 Study Design . . . 102
6.2.1 Survey Development . . . 104
6.2.2 Study Execution . . . 104
6.2.3 Analysis . . . 104
6.3 Understanding Prolific Crowdworkers from their Response of Survey Questions 105
6.3.1 General Demographics . . . 107
6.3.2 Technical Demographics . . . 107
6.3.3 Crowdworkers Demographics . . . 109
6.3.4 Financial Demographics . . . 109
6.3.5 Crowdworkers' Device Usage . . . 109
6.3.6 Crowdworkers Opinions . . . 117
6.4 Comparison of Overall Findings Between Prolific and Amazon MTurk . . . 123
6.4.1 General Demographics . . . 124
6.4.2 Owned Device Type . . . 124
6.4.3 Tool Usage . . . 124
6.4.4 Current device Usage Frequency for Management and Completion . . . 125
6.4.5 Desired Task Completion on Non-Desktop Devices . . . 125
6.4.6 Desired Task Management on Non-Desktop Devices . . . 126

6.4.7 Challenges with Device Use 126
6.4.8 Features related to Crowdwork Ecosystem 127
6.5 Clusters of crowdworkers in Prolific 127
6.5.1 Analysis 127
6.5.2 Result 130
6.5.3 Description of each PCA Component 130
6.5.4 Description of Clusters of Prolific Crowdworkers 140
6.6 Comparison of Clusters of Crowdworkers between MTurk and Prolific 151
6.7 Correlation of Diferent types of Demographics and Crowdworkers' Work Practices 152
6.7.1 Method 152
6.7.2 Correlation of General Demographics and Crowdworkers Work Practices 153
6.7.3 Correlation of Technical Factors and Crowdworkers Work Practices . 157
6.7.4 Correlation of Financial Factors and Crowdworkers Work Practices . 162
6.7.5 Correlation of Profession outside of Crowdwork and Crowdworkers Work Practices 162
6.8 Discussion 164
6.8.1 Prioritize Features in Platforms for Younger Adults 164
6.8.2 Tailor Task Complexity to Educational Backgrounds 165
6.8.3 Optimize Interfaces for Commonly Owned and Platform Specific Preferences 166
6.8.4 Support Tool Integration and Development 167
6.8.5 Tailor specific Tasks for Completion on Non-desktop Devices 167
6.8.6 Optimize Task Management fo Commonly mentioned Non-desktop devices 168
6.8.7 Implement Crowdwork Ecosystem Support Features 169
6.8.8 Region Specific Design for Platforms 169
6.9 Limitation 170

7 Incentives of Platform Owners , Potential Risk and Limitation 172

7.1 Financial Incentives of Platform Owners vs. Traits to Support Worker Flexibility 172

7.2 Potential Risk for Using Various Devices in Crowdwork 173
7.3 Limitation 173

8 Discussion **175**
8.1 Broader Implications in the Workflow of Society 175
8.1.1 Impact on Flexibility on Different Type of Profession 175
8.1.2 Informing Flexible and Supportive Work Policies 176
8.1.3 Economic Development and Innovation 176
8.2 Implications in the Future of AI supported Crowdworld 177
8.2.1 AI-Driven Workflow Optimization 177
8.2.2 AI-Driven Scripts and Tool Unification for Efficiency 178
8.2.3 Cross-Device AI Integration for Enhanced Task Management 178
8.2.4 Personalized AI Support Systems 178
8.2.5 AI-Powered Platform Insights 179
8.2.6 AI driven Solutions to Prevent Bots in Crowdwork 179
8.2.7 Enhanced Communication and Design through AI 180
8.2.8 AI Supported Equal Work Opportunity for All Crowdworkers 180
8.2.9 Ethical and Privacy Considerations 181

9 Conclusion **182**

Bibliography **185**

Appendix **212**
A Study Instrument For Understanding Crowdworkers Practices and Desire in MTurk 212
B Study Instrument For Understanding Crowdworkers Practices and Desire in Prolific 226

Vita **241**

List of Tables

2.1 List of four non-workstation devices studied in this thesis. Includes details about their advantages, disadvantages, and input/output characteristic. Also includes key citations regarding HCI research on these devices. 8

3.1 The taxonomy of characteristics and their associated values. Each characteristic was rated as contributing to a particular Mobile Usability Rating (MUR): ■ Good, ■ Fair, or ■ Poor. A * indicates that MUR is heavily context-dependent. 21

3.2 Usability values (MUR) for Good, Fair and Poor and usability percentage aggregated across seven characteristics across each HIT type. We order each HIT type by the Usability Percentage. 39

3.3 Results from Chi-Squared tests across task types. These tests compare the distribution of each HIT Type's distribution of usability values (MUR) to the other HITs in the set. Significance indicates a relatively distinct profile of usability. (*****: p<0.05; ******: p<0.01) 43

4.1 Demographics for the participants taking the study, less those that were removed for quality reasons 57

5.1 Top features of PCA Component 1 along with their loading factor within threshold absolute value 0.10 81

5.2 Top Features of of PCA Component 2 along with their loading factor within theshold absolute value 0.10 82

5.3 Each cluster along with their size and each two PCA component's mean value 82

5.4 Each cluster along with their size and each PCA component's median value . 82

5.5 All the significant features along with their p-value within threshold effect size of more than 0.3 across all three clusters of workers 87

6.1 Usage distribution of various crowdsourcing platforms mentioned by participants 111
6.2 Usage Distribution of Tools 112
6.3 Distribution of challenges mentioned by participants when using different devices for crowdwork 118
6.4 Top features of PCA Component 1 along with their loading factor within threshold absolute value 0.10 131
6.5 Top features of PCA Component 2 along with their loading factor within threshold absolute value 0.10 132
6.6 Top features of PCA Component 3 along with their loading factor within threshold absolute value 0.10 133
6.7 Top features of PCA Component 4 along with their loading factor within threshold absolute value 0.05 134
6.8 Top features of PCA Component 5 along with their loading factor within threshold absolute value 0.05 135
6.9 PCA Mean Values along with six clusters and their size 141
6.10 PCA Median Values along with six clusters and their size 141
6.11 All the significant features along with their p-value within threshold effect size of more than 0.5 across all six clusters of crowdworkers 154
6.12 Significant features that show the correlation between various demographics, technical factors, and professions outside of crowdwork with crowdworkers' device usage for work practices 154
6.13 No significant feature that shows the correlation between financial factors and crowdworkers' work practices 163

List of Figures

3.1 We present our taxonomy on HIT characteristics for mobile usability. We have organized the taxonomy according to Task and Interaction characteristics. The Poor, Fair, and Good legend map to the Mobile Usability Rating (MUR) more defined throughout this paper, with definitions of underlying values in Table 3.1 and results from an in-the-wild HITs demonstration in Section 5 (e.g. Figure 3.8). 16

3.2 Representation of themes in responses to Q24.2. 31

3.3 Histogram of task types labels that are currently completed or require additional support on smartphones. 31

3.4 Representation of labels for "Content Visibility" across the dataset of 519 HITs. 37

3.5 Representation of attributes for task characteristics: (a) Off-Site Knowledge Demands with 26.1% Present and 73.9% None, (b) Task Divisibility with 74.7% divisible, 21.8% Indivisible, and 3.4% Already Divided, (c) Task Size Volume with Low Volume with 81.2% High Volume and 18.8% Low Volume, and (d) Task Completion Time with 8.8% fast, 64.8% Normal, and 26.4% Slow. 37

3.6 Representation of attributes for interaction characteristics: (a) Multi-device Demands, (b) Prompt Modality, (c) Interaction Style. 39

3.7 Distribution of Good, Fair, and Poor across each HIT type. Here, we order by the least amount of total Poor usability values. This provides a ranking of Tasks, ordered by most to least usable according to our taxonomy. 41

3.8 Distribution of Good, Fair, and Poor attributes across all seven taxonomy characteristics binned by HIT type. 43

3.9 Three example HITs from our scraped corpus with their following Mobile Usability Rating (MUR) based on our taxonomy of characteristics (Figures 3.1 and 3.8). Each has a different Task Type and corresponding MUR ratings. The Sentiment Analysis / Text Classification HIT has an excellent MUR rating. In particular, the Interaction Style and low Task Size Volume is well-aligned for completing on a mobile phone. In contrast, the Information Finding HIT has a poor MUR rating. It requires switching away from the main HIT interface to a website, searching for URLs, and entering the answers with an Interaction Style of Open-ended+. In the middle, the Content Generation HIT can be performed on mobile phone, listening to short clips of pop-music from the same page, but still uses and Interaction Style of Open-ended+. 44

4.1 Frequency with which different devices are used to complete HITs. 59
4.2 Factors Influencing Crowdworkers' Choice of Device for Completing Tasks . . 59
4.3 Types of HITs crowdworkers are currently completing with different non-workstation devices. Data is from an open-response question. Percentages are based on the number of participants that have used each device to complete HITs. 60
4.4 Perceived Ease of Task Completion by Crowdworkers Utilizing Various Devices. 5 is high usability and 1 is low usability. 60
4.5 Crowdworkers' reaction to the impact of having their workstation break. . . 62
4.6 Crowdworkers' perceptions on the potential effectiveness of using an alternative device if their workstation breaks. 62
4.7 Frequency with which different devices are used to manage HITs. 62
4.8 Factors Influencing Crowdworkers' Choice of Device for Managing Tasks . . 64
4.9 "Perceived Ease of Task Management by Crowdworkers Utilizing Various Devices. 5 is high usability and 1 is low usability. 64
4.10 Predominant tools utilized by crowdworkers in our study 66
4.11 The number of tools each crowdworkers uses in our study 66

4.12 Types of HITs crowdworkers from our study are interested in completing with different non-workstation devices. Data is from an open-response question. Percentages are based on the number of participants that have used each device to complete HITs. 66
4.13 The percentage of crowdworkers from our study that want to perform HIT management tasks with different non-workstation devices. Data is from a close-ended question. Percentages are based on the number of participants that own each device. 67
4.14 The percentage of crowdworkers from our study that indicated various items when asked how they would use a magic wand to improve crowdwork on different devices. Data is from an open-response question. Percentages are based on the number of participants that own each device. 67

5.1 Scree Plot which shows the optimized number of PCA components 2 for our dataset 80
5.2 Elbow plot which shows the optimized number of clusters are 3 for our dataset 80
5.3 Siloutte Score Plot which shows the optimized number of clustering are 3 in our case 81
5.4 Visualization of all the significant features as shown in Table 5.5 within threshold effect size of more than 0.3 across all three clusters of workers . . . 87
5.5 Decision Tree which shows all unrolled features that differentiate clusters of crowdworkers 88

6.1 Visualization of distribution of residence of participants 106
6.2 Visualization of distribution of mostly mentioned cultural background of participants 108
6.3 Visualization of device usage frequency for day to day tasks outside of crowdwork 108
6.4 Visualization of agreement in interest in technology by participants 110
6.5 Visualization of no of tools used by participants 110
6.6 Visualization of no of years participants working as a crowdworker 111
6.7 Visualization of no of hours participants use to complete crowdwork 112

6.8 Visualization of no of hours participants use to manage crowdwork 113
6.9 Visualization of distribution of financial situation mentioned by participants 114
6.10 Visualization of distribution of profession outside of crowdwork 118
6.11 Visualization of distribution of current device usage for completing crowdwork 118
6.12 Visualization of distribution of current device usage for managing crowdwork 119
6.13 Visualization of importance of place for device usage in crowdwork 119
6.14 Visualization of importance of time for device usage in crowdwork 119
6.15 Visualization of frequency of device usage in an ideal world for completing crowdwork 120
6.16 High interest of completing five different types of tasks including survey, testing, transcription, classification and information finding across studied devices by participants 120
6.17 Visualization of frequency of device usage in an ideal world for managing crowdwork 121
6.18 High interest of managing four different types of tasks including finding and accepting, communication with requesters and workers as well as receiving notifications, across studied devices by participants 121
6.19 Visualization of importance of various factors when managing and completing crowdwork including time, place, device and variety of task according to participants 128
6.20 Visualization of importance of platform supported features related to completion and managment of crowdwork to participants 128
6.21 Visualization of importance of platform supported features realted to crowdwork ecosystem to participants 129
6.22 Scree Plot which shows the optimized number of PCA components 5 for our dataset 129
6.23 Elbow plot which shows the optimized number of clusters are 6 for our dataset 141
6.24 Siloutee score plot which shows the optimized number of clustering is 6 in our case 142

6.25 Visualization of all the significant features as shown in Table 6.11 within threshold effect size of more than 0.5 across all six clusters of crowdworkers . 149
6.26 Decision Tree which showsall unrolled features that differentiate clusters of crowdworkers 150
6.27 Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to gender and current smartphone use to manage crowdwork 155
6.28 Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to currently manage crowdwork 155
6.29 Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to manage crowdwork in ideal world 156
6.30 Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to complete crowdwork currently 156
6.31 Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using smartphone to complete crowdwork currently 158
6.32 Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to currently complete crowdwork 158

6.33 Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to complete crowdwork in ideal world 159
6.34 Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to currently manage crowdwork 159
6.35 Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to manage crowdwork in ideal world 160
6.36 Visualization of the significant feature a shown in Table 6.12 between technical factors and crowdworkers' work practice that have correlation related to tech interest and using smart speaker to complete crowdwork in ideal world . . . 160
6.37 Visualization of the significant fearure as shown in Table 6.12 that shows correlation between workers' profession outside of crowdwork and smart speaker use to manage crowdwork in ideal world 163

9.1 Overview of studied traits 184

Chapter 1

Introduction

1.1 Motivation

Crowdwork is a contemporary form of on-demand information work that involves the completion of independent tasks of varying complexity, difficulty, knowledge demands, and time constraints. These tasks, often referred to as Human Intelligence Tasks (HITs) , are created by *requesters* (i.e., the people who make work available) on crowdsourcing platforms such as Amazon Mechanical Turk *, Prolific † in order to find *crowdworkers* (i.e., the people who work on crowdsourcing platforms) who will complete work in exchange for monetary reward. With the rise of these platforms , crowdwork has become integral as digital professions in the 21st century with "tens of thousands of new workers" arriving on crowdsourcing platforms each year [39, 117].

Online surveys with residents in the United Kingdom and the European Union suggest crowdwork is widespread in nature, reporting that millions of citizens engage in crowdwork with a substantial percentage mentioning it as their full-time job [96, 97]. Further participation on these platforms is expected to grow globally at an annual rate of 26% [109] . Such growth is important as crowdwork has become a highly instrumental research tool [20, 140], particularly in the areas of HCI [143, 185] and ML/AI [149, 1].

*https://www.mturk.com/
†https://www.prolific.com/

Crowdworkers are drawn to crowdwork for a myriad of reasons [116, 175], notably, one of the reasons crowdworkers are attracted to crowdsourcing platforms because of flexibility it promotes. This includes flexibility of when work is completed such as temporal fleixbility [236, 55, 120], where it is completed [241, 42], and how it is completed [42, 236] such as spatial fleixbility. Flexibility in work has been demonstrated to offer numerous advantages for traditional employees. For example, greater flexibility can lead to enhanced productivity, increased job satisfaction, and better mental and physical health [12, 158]. Additionally, flexible work arrangements allow traditional employees to more effectively balance work with their non-work commitments [195]. Individuals who have the freedom to work anytime and anywhere typically experience lower levels of work-life conflict, report less stress, and are more content with their work setups [59, 178].

There is active research happening focusing on the intricacies of temporal and spatial flexibility in crowdwork. A prime example of this focus is the work by Lascau et al., which delves into how crowdsourcing platforms might simultaneously restrict and enhance crowdworkers' temporal flexibility, particularly for those who value the ability to set their own schedules [120]. This exploration extends to the impact of platform policies and job designs on spatial and task flexibility. Researchers are keenly analyzing the repercussions of practices such as the withholding of critical task information and the specific ways tasks are allocated. Such practices, they find, can significantly disrupt workers' planning capabilities and efficiency [42]. This body of research aims to paint a comprehensive picture of how these elements of flexibility interplay to shape the crowdwork landscape, striving for insights that could lead to more worker-friendly platform designs.

Despite these advancements, the current design of crowdsourcing platforms often limits the flexibility of crowdworkers. A substantial portion of crowdwork management and completion is still primarily conducted on workstation or desktop computers [229, 187, 212]. This reliance on stationary devices restricts workers from performing tasks when, where, and how they prefer, thereby limiting their flexibility.

To address this limitation, the overall goal of this thesis is to better support the flexibility of crowdworkers. Prior research suggests that incorporating a variety of device types in the workflow plays a crucial role in enhancing crowdworkers' flexibility [89]. Despite evidence that crowdwork can be performed on diverse device types, the question remains: why is crowdwork still predominantly desktop/ workstation-centric?

While past studies have demonstrated that crowdwork can be done on non-desktop devices [66, 2, 87, 44], they have not addressed the need for tailored workflows for each device type. Each device type requires a unique workflow to accommodate its specific traits. The standard workflow for crowdwork includes stages such as managing tasks (including finding, accepting, creating, and listening to catchers, and communication with requesters and workers) as well as completing different types of tasks such as information finding and classification, etc. [229, 159, 58]. I hypothesize that different devices will exhibit unique traits for these stages.

In this dissertation, I explore what those traits are. By understanding the specific needs and capabilities of various devices, we can develop tailored workflows and interfaces to best support each device type. This research lays the groundwork for future work to create more flexible and accommodating crowdsourcing platforms, ultimately enhancing the experience and efficiency of crowdworkers. I next outline research goals I will address in this thesis.

1.1.1 Research Goal

In this thesis, I focus on the following research goals:

1. I first examine task completion on smartphones, focusing on task usability, given that smartphones are the most commonly owned and used device type.

2. Expanding the list of non-workstation devices, I then explore crowdworkers' current task management and task completion practices and expectations when working on smartphones, tablets, speakers, and smartwatches.

3. With a broad understanding of crowdworkers' current and desired practices and expectations across different devices, I aim to identify systematic differences among crowdworkers.

4. All previous findings are structured around one crowdsourcing platform, Amazon MTurk. Therefore, I extend this work to another prominent crowdsourcing platform, Prolific, to

quantify the qualitative results from MTurk and compare the results between MTurk and Prolific.

This comprehensive approach aims to uncover the traits that contribute to the flexibility of crowdworkers, ultimately striving to inform the design of more adaptable and worker-friendly crowdsourcing platforms.

1.2 Research Contributions

This thesis provides several key takeaways that have significant implications for the future of the crowdwork field, especially in the context of crowdwork being a critical component in training and evaluating AI/ML models:

The thesis underscores the importance of designing crowdsourcing tasks and platforms that are highly accessible and usable across a variety of devices. For example, our findings show that crowdworkers show a preference for completing and managing their tasks primarily through smartphones and tablets other than desktop or workstation, while smart speakers and smartwatches are preferred for specific quick task management. This device versatility ensures that crowdworkers can engage with tasks using their preferred devices, such as using a tablet for detailed survey completion or a smartwatch for receiving task notifications, increasing their flexibility and ability to complete and manage work effectively.

The thesis highlights the necessity of personalized crowdsourcing platforms that can adapt to the unique work practices and preferences of individual crowdworkers. By employing detailed quantitative analysis and ML-based modeling, the research uncovers systematic differences in crowdworkers' practices at different stages of their work. For instance, majority crowdworkers in MTurk prefer to start their tasks on a desktop , however they are open to managing tasks such as finding and accepting tasks on smartphone and tablet for quick updates or adjustments on the go. Such personalization can lead to more efficient workflows and higher satisfaction among workers, ensuring that the platforms can provide a tailored experience that meets individual needs.

The research provides robust platform-driven insights for understanding crowdworker behaviors and preferences. Through comprehensive qualitative and quantitative analyses, the thesis offers a deep understanding of the current practices and expectations of crowdworkers across various device types, including workstations, smartphones, tablets, smart speakers, and smartwatches. For example, although most of workers on both MTurk and Prolific workers prefer desktop , MTurk workers mostly open to smartphone and tablet; wheras in Prolific there are group of workers who prefer using very non traditional devices such as smart speaker, smart watch. These insights are crucial for optimizing platform design and functionality, ensuring alignment with the evolving needs of the crowdwork community.

As crowdwork continues to be a foundational element for training and evaluating AI/ML models, the improved design of crowdsourcing platforms and tasks will enhance the quality and reliability of the data collected. By providing detailed quantitative and qualitative analysis, the research offers a framework for designing platforms that can better support the flexibility of of crowdworkers.

Chapter 2

Background and Related Works

This section provides the background information necessary to supplement the understanding of the work presented in this dissertation. We begin by discussing the prevalent nature of crowdwork, emphasizing its desktop-centric practice and its impact on general flexibility. We first provide a thorough review of HCI studies focused on four key non-workstation devices: smartphones, tablets, smart speakers, and smartwatches. Following this, we delve into existing research to understand how crowdworkers utilize these devices to complete and manage tasks. Finally, we conclude the relation of existing work with the thesis which focuses to understand the factors and characteristics to support crowdworkers flexbility particiularity related to device usage.

2.1 The Desktop-Centric Nature of Crowdwork and Workstation's Relationship with Flexibility

Crowdwork is predominantly a desktop-centric practice, with the vast majority of crowdworkers relying on workstation or desktop computers as their primary devices for work-related activities. Quantitative and qualitative studies similarly report that, as in many other information work professions, crowdworkers recognize the advantages of using these devices over other alternatives [89, 229].

The utility of workstation computers in crowdwork primarily stems from their ability to support the nature of on-demand work, where tasks must be captured and completed efficiently. Workstations provide larger screens [89], which facilitate easier navigation and management of multiple tasks simultaneously. This is crucial for handling the diverse activities involved in crowdwork, such as data entry, content moderation, and complex problem-solving tasks. The enhanced processing power of workstations allows for smoother handling of resource-intensive applications and data. Additionally, workstations offer robust multitasking capabilities, enabling crowdworkers to run multiple programs and browser tabs without significant performance degradation.

Moreover, the reliance on workstation computers is evident in the community-driven infrastructure that supports crowdwork. Platforms for reviewing tasks or requesters, such as Turkopticon [101], and forums for connecting with other crowdworkers, such as TurkerView [187], are primarily designed for desktop use. These platforms provide critical resources for crowdworkers to enhance their efficiency and effectiveness, reinforcing the necessity of workstation computers in their daily workflow. Desktop interfaces for these platforms often include detailed dashboards, advanced search capabilities, and comprehensive review and feedback systems. Additionally, the integration of various browser extensions and plugins, which crowdworkers frequently use to automate and streamline their tasks, is generally more effective on desktop systems.

This reliance on workstation computers ties people to their workstations, significantly constraining their flexibility. Being tethered to a stationary setup limits the ability to work from various locations and adapt to different environments, which can affect work-life balance and overall mobility [183, 80]. The necessity to remain at a desk for extended periods can also lead to physical discomfort and health issues, further impacting productivity and well-being [25].

Table 2.1: List of four non-workstation devices studied in this thesis. Includes details about their advantages, disadvantages, and input/output characteristic. Also includes key citations regarding HCI research on these devices.

Device	**Advantages**	**Disadvantages**	**Input/Output Characteristics**	**Citations**
Workstation	More comfortable over long durations. Easier to enter text (i.e., has a physical keyboard).	May be tied to a single location (i.e., desktops) or difficult to use in transit between locations.	Mouse, keyboard, large screen size, speakers or headphones	[104, 57]
Smartphone	Ubiquitous Internet connection. Usually remains with the individual at all times.	Very small screen. Difficult to enter textual data (i.e., uses a small virtual keyboard)	Touch-based input, camera, microphone, small screen size, speakers or headphones	[104, 166, 44]
Tablet	Large touchscreen display. Easier to carry around on use when on the go than a laptop.	Difficult to enter textual data (i.e., uses a virtual keyboard), mediocre camera, mediocre microphone	Touch-based input, camera, microphone, medium screen size, speakers or headphones	[166, 159]
Smart Speaker	Can access without needing to physically interact. AI-supported processing of input.	Can be unreliable due to ambient noise. No screen or keyboard.	Microphone, speaker	[211, 88]
Smartwatch	Often accompanied by a ubiquitous Internet connection. Remains with the individual at all times. Easy to see and track notifications.	Extremely limited screen size. Not feasible to have a virtual keyboard.	Touch-based input, mediocre camera, mediocre microphone, biometric sensors, speakers	[56, 163, 157]

2.1.1 Non-Workstation Device Usage in HCI and Its Relationship with Flexibility

Today's everyday life including work life is characterized by the spread of non-workstation devices ranging from wireless devices such as smartphones, tablets, smartwatches, and smart speakers. The adoption and usage of these non-workstation devices are rapidly growing in prominence. Table 2.1 provides a broad overview of both workstation and non-workstation devices. By presenting workstation devices as a baseline, it facilitates a clearer comprehension of the benefits and drawbacks associated with four non-workstation devices studied in this work. We then discuss related literature on non-workstation devices' ability to improve crowdworkers' flexibility.

Smartphones

Smartphones are used every day by a majority of people. For example, Pew Research Center reports that 85% of American adults own a smartphone [24]. Smartphones have been mainly used in work for doing brief tasks (e.g. viewing calendar, taking small notes) [165], sending and receiving e-mails [144, 165], communication (e.g., text messages, emails, and phone calls) [147]. Other arguments for supporting flexibility associated with smartphones are increased job performance, productivity, enthusiasm to achieve the organization's goals and missions, flexibility in work schedules and improved trust among organizational members, responsiveness, and the availability of real-time information [131, 147, 129]. Beyond smartphones, people are also interested to use other non-workstation-based devices (e.g. tablets, smart speakers, and smartwatches).

Tablets

Tablets are now widely used by people of all ages. According to U.S. Census, Tablet ownership lagged behind smartphones, desktops, and laptops, with 63 percent of households owning them [139]. Prior work has shown that tablets are used differently in different work settings. This device type is used in general education and learning for searching the web, checking email, for reading work materials [227, 166]. It not only assists and improves learning, comprehension, and document management, but they also make courses more fascinating and pleasant by providing the flexibility to use it anywhere, anytime as a learning medium [230, 41, 202, 52, 93, 201]. Tablets have been also found to be useful in corporate environments for work-related meetings, reading emails, multi-tasking, organizing tasks, and displaying and analyzing visual data [60]. Besides that, tablets are also helpful in healthcare organizations for analyzing MRI brain scans as well as a preferred computing device for healthcare professionals for fast access to information [60, 148].

Smart Speakers

Smart speakers are increasingly adopted by people. According to a consumer report, more than one-quarter of households own a smart speaker in the U.S. [115]. A study of 18 households' found that adults and children use smart speakers for different activities [62]. Adults most commonly use smart speakers to provide automation flexibility (e.g. commands for controlling other devices or IoT systems such as smart lights, thermostats, and cameras) and music commands (e.g. playing music based on the name of a song, artist, or album). People tend to use smart speakers for efficient information-seeking tasks in different ways for example as a type of information kiosk that allowed students to ask about events and class schedules with voice control [191], to answer inquiries in a public slum area [169], health-related information seeking [18]. This device has type has also been used for health-related activities such as monitoring fitness using audio signals [233], and supporting women with metastatic breast cancer [177].

Smartwatches

Consumers have shown strong interest in smartwatches as well. Most activity on smartwatches currently includes timekeeping, notification, activity tracking, and as an input device for different applications (e.g. apps related to health and fitness, navigation (e.g. GPS), social media), for fitness monitoring by providing smart exercise guidance [188, 77, 174]. Besides that, the use of smartwatches has been reported by healthcare practitioners. For example, for medical education, patient care, and community outreach[154]. Bernaerts et al. employed smartwatches at the workplace to physically and digitally lock and unlock doors, obtain room information, and issue virtual knocks with an app running on the watch. This provides great flexibility to the employees since they no longer need to carry keys or key cards for their office buildings [13].

2.2 Crowdwork on Non-Desktop Devices

Prior research highlighted the potential of mobile devices in crowdsourcing. TxtEagle [47] and mClerk [78], deployed in Kenya and India respectively, enabled individuals to complete tasks via SMS text messages. These tasks included language translation, market research, audio transcription, and low-cost image classification, demonstrating that simple mobile phones could be leveraged for crowdwork without the need for advanced computing power.

Building on this foundation, Narula et al. developed MobileWorks, a smartphone-based platform for optical character recognition tasks [152]. Similarly, Yan et al. introduced an iOS application, mCrowd, for posting and submitting sensor-related crowdsourcing tasks using smartphones [234]. These applications showcased the ability of smartphones to handle more complex and varied crowdwork tasks.

Voice-based interactions on smartphone have also been explored as a means to enhance flexibility in crowdwork. Vashistha et al. demonstrated Respeak, a voice-driven system for assisted transcription on mobile devices, highlighting how voice interactions can simplify task completion on the go [217]. Google's Crowdsource app further exemplifies this trend, enabling users to complete small HITs like image transcription, translation, and handwriting recognition directly on their smartphones [26]. Chopra et al. developed an Android application for digitizing handwritten Marathi/Hindi words, demonstrating the use of mobile devices for specific linguistic tasks [27]. Williams et al. [229] and Newlands et al. [159] investigate the use of smartphones for crowdwork management.

In addition to smartphone, there has been more research which emphasized the increasingly multi-device landscape of crowdwork. Goncalves et al. [67, 70] showed that tablets can be successfully used for completing tasks targeting individuals in public spaces with idle time, while Hosio et al. [90] demonstrated the feasibility of a situated crowdsourcing system for completing tasks. Hettiachchi et al. proposed several approaches for using smart speakers to complete HITs, and their system, Crowd Tasker, delivered HITs via a digital voice assistant. Despite initial usability issues, participants found that using a smart speaker allowed them to multitask more effectively, providing greater flexibility for HIT completion [88, 87]. In the realm of wearable technology, Acer et al. introduced a method that integrated mobile postal workers into crowdsourcing tasks using a wearable smartwatch app. This system directed workers to relevant locations via enhanced routes through notifications, improving response rates and task accuracy [2]. This study highlighted the potential of wearable devices to augment the efficiency and precision of crowdwork tasks. In their paper, Calacci et al. showed how smartwatches can be used to delegate and manage complex tasks to crowd workers, and how this formalizes patterns of management in crowd work [21].

2.3 Relation to This Work

Previous research demonstrates the use of various devices in crowdwork and their impact on work flexibility. Work flexibility has been praised for its benefits, such as increased productivity and job satisfaction [12, 158]. Prior studies also suggests that incorporating a variety of device types into the workflow is crucial for enhancing crowdworker flexibility [89]. However, each device type requires a tailored workflow. Standard workflows for crowdworkers include stages such as managing and completing different types of tasks. I hypothesize that the factors and characteristics of task completion and management may vary to support tailored workflows for different device types.

To my knowledge, my work is the first to enhance the flexibility of crowdworkers by examining the use of non-workstation devices in their workflows. Given that smartphones are the most commonly used device type as shown in Section 2.2 , my first objective is to understand the characteristics of task completion and the factors affecting this process on smartphones. Expanding beyond smartphones, I then seek to explore crowdworkers' current practices and expectations when using other non-workstation devices, including tablets, smart speakers, and smartwatches. With a broad understanding of crowdworkers' practices and expectations across these different devices, I identify systematic differences among crowdworkers. This will help develop customizable support tailored to individual needs. Finally, as previous findings were primarily based on Amazon Mechanical Turk, I also investigate whether the platform itself influences the explored questions by validating the generalizability of the findings using another prominent crowdsourcing platform, Prolific.

Chapter 3

Understanding the traits to support crowdwork completion on mobile devices*

3.1 Introduction

Studies of crowdwork generally describe the digital profession as one centered around the workstation computer. In crowdwork specifically, much of the utility afforded by workstation or desktop computers stems from their ability to support the nature of on-demand work in which tasks must be captured and completed efficiently. Specific motivations for desktop-centric work practices in crowdwork include screen-size demands [89], limitations of productivity (e.g., HIT finding [229]), and general ease in completing administrative tasks related to crowdwork (e.g., reviewing requesters [187, 212]). A variety of efforts ranging from individuals apps (e.g., Respeak [217]) to full-blown platforms (e.g., Google's *Crowdsource* [26]) independently have facilitated crowdsourced work experiences that are designed to be completed on smartphones. In contrast to these prior contexts, modern crowdsourcing platforms allow requesters to build task interfaces themselves, providing few to no formatting constraints on way in which a task's question its posed or or the way in which a task's response is collected. Each HIT is

*This chapter is adopted from my publication: Dutta, Senjuti, Rhema Linder, Doug Lowe, Richard Rosenbalm, Anastasia Kuzminykh, and Alex C. Williams. "Mobilizing crowdwork: A systematic assessment of the mobile usability of hits." In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1-20. 2022.

structurally unique from the next, and for this reason, it remains unclear how crowdwork, more generally, can be translated to a more mobile context.

The notion of mobilizing crowdwork hinges on understanding the characteristics that shape the usability of a HIT's interface on a mobile device. *Usability* is an aspect of software design that impacts the satisfaction and effectiveness of its users in performing tasks and activities [31]. Usability can be evaluated with design processes from criteria, empirical tests in laboratories, and is contextually valid based on the goals of stakeholders [176]. Prior studies have shown that the usability of mobile devices and the usability of traditional workstations are distinguished by having different constraints, such as a necessity for small screen size and reduced computational power [239]. Therefore, HITs that have high usability – e.g. are usable – for workstations may lack mobile usability. By better characterizing the distinctions that separate mobile and workstation usability in crowdwork, researchers, requesters, and crowdwork platforms will be able to understand pathways for making work more accessible on mobile devices.

To address this issue, we present a taxonomy of characteristics that defines the mobile usability of HITs for smartphone devices.[†]. The taxonomy is developed based on findings from a study of three consecutive steps. In Step 1, we first contextualize our work by using a workshop to generate and discuss characteristics in relation to relevant literature. Then, we apply this derived knowledge towards accumulated HIT suitability research which is focused towards the taxonomy of characteristics that relates to mobile suitability. In Step 2, we issue and analyze an online survey that provides empirical support to the developed taxonomy with added nuance from the perspective of Mechanical Turk workers. Finally, in Step 3, we use the taxonomy to evaluate the mobile usability of 519 HITs on Amazon Mechanical Turk.

[†]This chapter is adopted from my publication: Dutta S, Linder R, Lowe D, Rosenbalm R, Kuzminykh A, Williams AC. Mobilizing crowdwork: A systematic assessment of the mobile usability of hits. InProceedings of the 2022 CHI Conference on Human Factors in Computing Systems 2022 Apr 29 (pp. 1-20).

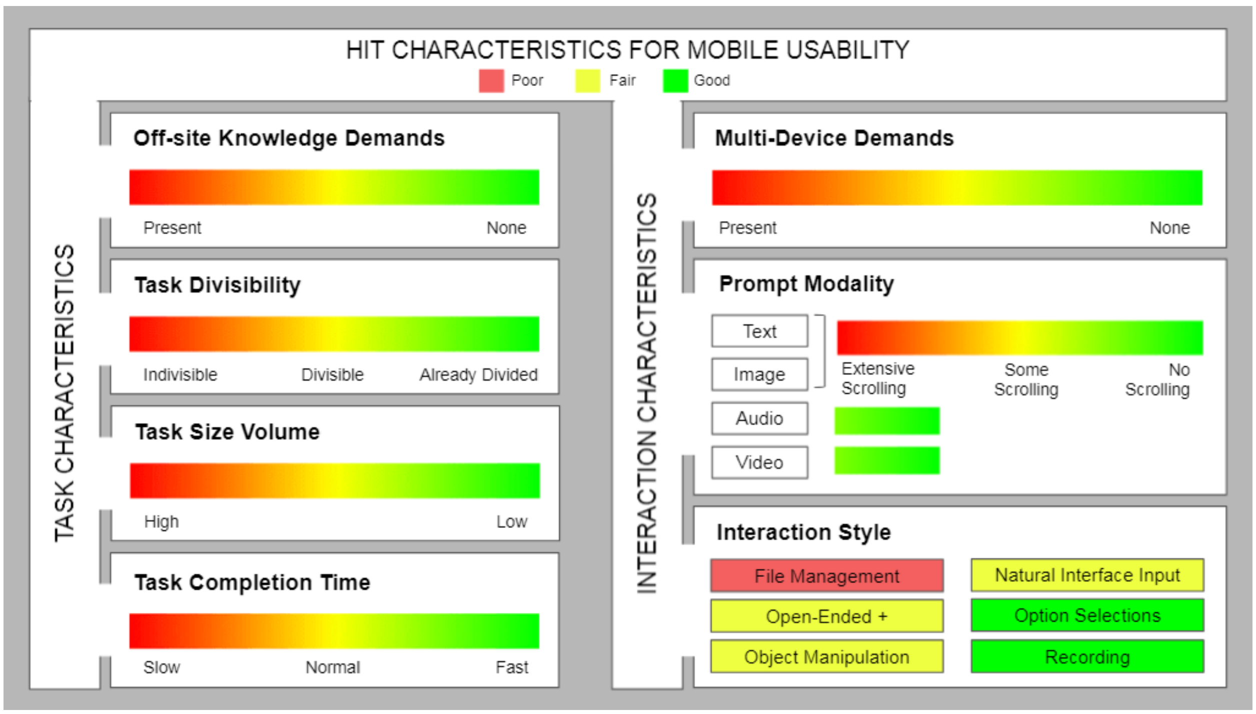


Figure 3.1: We present our taxonomy on HIT characteristics for mobile usability. We have organized the taxonomy according to Task and Interaction characteristics. The Poor, Fair, and Good legend map to the Mobile Usability Rating (MUR) more defined throughout this paper, with definitions of underlying values in Table 3.1 and results from an in-the-wild HITs demonstration in Section 5 (e.g. Figure 3.8).

3.2 Taxonomy Generation - Step 1: Targeted Literature Analysis

The goal of our research is to understand the characteristics of HITs that contribute to their usability on mobile smartphone devices. To better understand these characteristics, we designed and employed a process heavily inspired by the Nominal Group Technique [32] that allow us to generate a taxonomy through brainstorming, ideation, discussion, and deliberation. In this section, we describe this process and conclude by presenting the final taxonomy of characteristics.

3.2.1 Procedure

We designed a procedure that centers around the Nominal Group Technique [32], a group-based decision making workshop method that aims to facilitate the generation of ideas from a small group of qualified experts or professionals. Our research team includes experts in crowdsourcing that have acted as requesters and workers and have a background in HCI. This justifies our process for proposing areas of interest for mobile usability. The technique's four-step process involves 1) generating ideas, 2) recording ideas, 3) discussing ideas, and 4) deliberating ideas. Methodological analyses of the Nominal Group Technique (NGT) specifically reinforce its use as a tool for computer-mediated ideation and brainstorming [50]. The procedure has been used in several contexts within human-computer interaction settings, including summaries of workshop activities [83], understanding system requirements [40], and improving computer-mediated decision-making [119]. Here, we leverage this technique to facilitate group-based decisions about the development of our taxonomy while drawing inspiration from related work.

Overview of Procedural Execution

Three members of the research team participated in the procedure. Two researchers currently hold a PhD in Computer Science while the third researcher is a PhD student in Computer Science. All three researchers have significant expertise and published research papers in crowdsourcing and crowdwork.

Before beginning the NGT procedure, each member of the research team independently conducted an review of literature toward the goal of searching for issues or concerns related not only to completing HITs on mobile devices, but also general usability related to mobile devices. Specific guidelines, suggestions, or taxonomies for designing HITs for mobile are not discovered during this preparatory phase of research. With the familiar literature in mind, our research goal shifted to summarizing research as characteristics, rather than to discovering or creating novel conceptualizations of what contributed to mobile usability on smartphone devices. In support of this goal, we engaged in a four-step, NGT-inspired process:

- **Stage 1. Idea Generation.** Our procedure began by assigning one researcher as the "moderator" who was responsible for facilitating the experience. The moderator asked each researcher, including themselves, to work silently and independently toward the goal of answering the following question:

  "What characteristics relate to a human intelligence task being completed on a mobile smartphone?"

  The overarching goal of beginning independently stems directly from the Nominal Group Technique which suggests that the quality and quantity of ideas that are produced [84, 146].

- **Stage 2. Idea Recording.** The moderator provided each researcher with a document (i.e., a Google Doc) in which they were required to record characteristics related to HIT mobility. Researchers were encouraged to provide a list of characteristics they perceived to be comprehensive and motivated by at least one prior work. Upon completion, researchers submitted the completed document to the moderator.

- **Stage 3. Idea Discussion.** After receiving each of their completed documents, the moderator organized a videoconferencing meeting in which they instructed each researcher to share and discuss the contents of their document. The primary goal of this stage was to ensure that members of the research team were given an opportunity to convey the importance of their documented characteristics. Each researcher took turns providing and receiving feedback, asking for clarifications, and producing new inquiries based on the discussion.

- **Stage 4. Idea Deliberation.** Following the presentation of ideas, the research team engaged in a deliberative process toward the goal of arriving at a consensus of ideas, which was used as the basis for developing the taxonomy of characteristics. As the research team developed more focused research questions, heuristics for including or excluding characteristics from the taxonomy emerged naturally. The inclusion of a specific characteristic was primarily motivated by referencing prior literature that supported its relevance toward answering the question defined in Stage 1. In contrast, a characteristic was excluded from consideration if one of the following conditions held: (1) related to a device, rather than an aspect of a HIT that could be configured by a requester, (2) related to aspects of "meta-work", such as finding or managing HITs, or (3) a characteristic that could not be easily assessed through visual or manual inspection (e.g., in a screenshot of the HIT). The research team concluded Stage 4 by arriving at a consensus set of ideas that underscore the mobile usability requirements of HITs. This included assigning a Mobile Usability Rating (MUR) of Good, Fair, or Poor for each characteristic's value per related work.

An important consideration for the design of this study is that our research team was required to operate in a remote and distributed fashion due to the on-going COVID-19 pandemic. While prior studies have utilized NGT to facilitate real-time brainstorming sessions that often take place in-person (e.g., [83]), our procedure took place entirely in a computer-mediated fashion using email communication and videoconferencing software (e.g., Zoom).

3.2.2 Findings

The first three stages of the NGT-inspired procedure resulted in a set of initial "ideas" for summarizing characteristics from literature. For clarity, we henceforth refer to the generated "ideas" simply as our "taxonomy of characteristics". Through the Idea Generation and Recording phases (*Stage 1 and 2*), a total of 20 non-unique characteristics were developed independently and suggested to the moderating researcher. The number of characteristics contributed by each researcher ranged from three to 11. Each characteristic was presented during the Idea Discussion phase (*Stage 3*). During this phase, a total of 8 suggested characteristics were identified as duplicates (i.e., reported by more than one researcher), leaving a total of 12 unique candidate characteristics for consideration in the taxonomy. During the Idea Deliberation phase (*Stage 4*), we developed heuristics for focusing characteristics as practical for assessments. This refined and modified how our summaries of characteristics in literature as attributes of these characteristics were more deeply and narrowly defined. The process of narrowing was guided by literature references that we subsequently used to fortify our characteristics with practical justification.

The NGT procedure concluded with a set of 7 characteristics of HITs that relate to their suitability for completion on mobile devices. Our taxonomy of characteristics is divided between two types: (1) *Task Characteristics* and (2) *Interaction Characteristics*. All possible values for each characteristic are driven by examples that arose from prior literature discussed throughout the later stages of the NGT procedure. Each possible value is mapped to a MUR value of "Good", "Fair", and "Poor" usability based on hypothetical contexts that arise through discussion as well as observed contexts reported in the literature. Importantly, each characteristic in the taxonomy is intended to assess a particular aspect of a HIT's suitability for use on smartphone devices. To assess the overall mobile HIT usability, each of these characteristics can be considered together. The set of characteristics and their associated values are shown in Table 3.1.

Task Characteristics

We identified a total of four characteristics that relate to a HIT's task design (i.e., the structural representation of task-related information). Task design is well-studied aspect of crowdsourcing that is concerned with the *efficiency* of crowdsourced tasks [123]. We draw on the following characteristics for our taxonomy:

1. *Task Completion Time* describes the amount of time required to complete a HIT. Prior work suggests that crowdworkers and information workers alike have an interest in using

Table 3.1: The taxonomy of characteristics and their associated values. Each characteristic was rated as contributing to a particular Mobile Usability Rating (MUR): ■ Good, ■ Fair, or ■ Poor. A * indicates that MUR is heavily context-dependent.

	Characteristic Name	MUR	Values	Description	Refs
Task Characteristics	Off-Site	■ (Good)	None	HITs that do not require information beyond the task interface.	[73, 240, 15]
	Knowledge Demands	■ (Poor)	Present	HITs that require information beyond the task interface.	[73, 240, 15]
	Task Divisibility	■ (Good)	Already Divided	HITs that are already divided into small tasks.	[225]
		■ (Fair)	Divisible	HITs that have clear boundaries for division into smaller tasks.	[91]
		■ (Poor)	Indivisible	HITs that lack clear boundaries for division into smaller tasks.	[231]
	Task Size Volume	■ (Good)	Low	HITs with content that includes two pieces of media or less.	[121]
		■ (Poor)	High	HITs with content that includes more than two pieces of media.	[71]
	Task Completion Time	■ (Good)	Fast	HITs that require 15 seconds or less to complete.	[14]
		■ (Fair)	Normal	HITs that require 15 to 60 seconds to complete.	[122]
		■ (Poor)	Slow	HITs that require more than 60 seconds to complete.	[98]
Interaction Characteristics	Multi-Device Demands	■ (Good)	None	HITs that do not require the use of multiple devices to complete.	[182, 89]
		■ (Poor)	Present	HITs that require the use of multiple devices to complete.	[182, 89]
	Prompt Modality	■ (Good)	Video	HITs that involve annotating and/or understanding video.	[8]
		■ (Good)	Audio	HITs that involve annotating and/or understanding audio.	[8]
		■ (Fair)	Image*	HITs that involve annotating and/or understanding images.	[8]
		■ (Fair)	Text*	HITs that require annotation and/or understanding text.	[8]
	Interaction Style	■ (Good)	Option Selections	HITs that require selecting a set of options (e.g. radio button).	[200]
		■ (Good)	Recording	HITs involving audio and video authoring.	[142]
		■ (Fair)	Open-ended+	HITs that require using fill-in-the-blanks or free-form text-entry.	[200]
		■ (Fair)	Object Manipulation	HITs that involve direct manipulation (e.g., bounding box).	[200]
		■ (Fair)	Natural Interface Input	HITs that involve unconventional input (e.g., gestures)	[200]
		■ (Poor)	File Management	HITs that require manipulating or uploading files.	[113]

smartphones only briefly for activities that can be completed quickly [8, 98, 229, 228]. The assumption is that faster tasks are preferred for mobile contexts.

2. *Task Divisibility* refers to the notion that a HIT can be broken down into smaller subtasks. Prior studies suggest that translating macrotasks to microtask counterparts increases the usability of these tasks on mobile devices [100, 228]. Similar characteristics had been suggested and discussed during Stages 2 and 3 (e.g., *Task Size Steps*), but were eliminated due to the difficulty associated from assessing its presence (e.g., from a screenshot).

3. *Task Size Volume* describes the information content within a HIT's task interface. Previous research indicates that larger task size volume negatively impacts crowdworkers' productivity [128].

4. *Off-Site Knowledge Demands* characterizes the underlying need to navigate away from a HIT's primary task interface (e.g., to find information on another webpage or make use of another web resource) in order to successfully complete it [73, 240, 15], an activity that is generally recognized as inefficient on mobile devices [184].

Interaction Characteristics

Alongside our four *Task Characteristics*, we identified a total of three characteristics that relate to a HIT's *interaction design* (i.e., the ways in which a task is posed and a crowdworker must complete it). The specific characteristics include:

1. *Multi-Device Demands* describes a HIT's underlying need for the use of multiple devices in order to complete it. Several recent studies on multi-device experiences in crowdwork suggest that the use of multiple devices is becoming increasingly more common [89].

2. *Prompt Modality* refers to the type of media that crowdworkers are prompted with and required to interface with (e.g., annotate, classify, etc). We draw directly from Peng et al. [170] to motive this characteristic's inclusion as it highlights four types of prompts.

3. *Interaction Style* details how a HIT requires crowdworkers to engage with it, whether it be through free-form text to complex natural language. We, again, chose to simplify

this characteristic toward the goal of observing interaction styles that can be assessed visually (e.g., with a screenshot) [200].

Thus, we have addressed RQ1 by developing characteristics in our taxonomy that are relevant to mobile HIT usability. We group the relevant usability characteristics in two sets *Task Characteristics* and *Interaction Characteristics* in Table 3.1.

3.3 Taxonomy Support - Step 2: Mechanical Turk Support Survey

Observations from our NGT-inspired study provided us with a taxonomy that describes the characteristics that contribute to a HIT's usability for completion on mobile devices. To better understand whether these characteristics capture the mobile usability requirements of HITs in reality, we draw from data collected from an online survey aimed at assessing mobile crowdwork through the lens of various mobile devices (e.g., smartphones, smartwatches, smart speakers).

3.3.1 Method: Online Survey

The original survey design was motivated by a broader research project aimed to understand the challenges and opportunities of engaging with crowdwork on mobile devices. The IRB approved survey includes 43 questions across five sections and is available as supplemental material[‡]. Though the research questions for this study were not specifically related to understanding the mobile usability requirements of HITs, several survey questions focused specifically on understanding the types of HITs that are suitable for mobile devices. We therefore conducted a targeted analysis of relevant questions to support the outlined taxonomy.

Survey Design

The survey began by inquiring about participants' personal demographics (i.e., age, gender, education) and their work experience on Amazon Mechanical Turk (e.g., completed HITs,

[‡]Please see the Supplemental Material section of Precision Conference.

current work hours, HIT approval rate). Thereafter, the survey was split into four sections collectively aimed at understanding current and desired mobile work practices in crowdwork. Our analysis specifically draws on three questions from two sections of the survey which is attached in the Appendix A:

1. *Section 2. Understanding HIT Completion.* This section includes multiple-choice and open-ended questions about the frequency and scenarios that crowdworkers utilize devices to complete HITs on MTurk. It also includes multiple-choice and open-ended questions about the frequency and scenarios that crowdworkers utilize devices to complete HITs on MTurk. From this section, we specifically analyze responses to the following questions:

    - Q15.1. *Please briefly describe what types of HITs you currently try to complete on your smartphone.*
    - Q17.2. *For your MTurk work, what types of HITs would you like to see better supported on the smartphone?*

2. *Section 5. The Magic Wand.* This section asks participants to consider a scenario in which they have a magic wand that allows them "to change whatever you'd like to change about work on Amazon Mechanical Turk to work on the platform how you want to work". From this section, we analyze responses to the following question:

    - Q24.2. *How would you use the magic wand to make managing and performing HITs better for a smartphone? What would you change? Why?*

Recruitment and Remuneration

We recruited a total of 111 participants for the study by deploying the survey to MTurk. To ensure participants' data was both reliable and motivated by experience with crowdwork, we employed a HIT qualification that required participants to have completed at least 10,000 HITs and have an a minimum acceptance rate of 98.0%. Prior research has found that crowdworkers on Amazon Mechanical Turk may multitask when payment is too low. We therefore chose to reward participants with $5.00 USD as it both ensures they are paid fairly and feel more comfortable devoting their complete attention to our HIT [229]. One participant reported that they do not own a smartphone, and six participants demonstrated spamming behavior in their survey responses. We chose to remove these seven participants, thus limiting our analysis to 104 participants.

Analysis Methods

All three survey questions relevant to the study at hand collected open-ended responses from participants. We conducted top-down coding analysis of two responses (Q15.1 and 17.1) , in which responses are categorized into pre-existing codes. We coded the third question (Q24.2) with a bottom-up approach, creating codes to find themes [17].

To analyze the responses responses to questions Q15.1 and 17.1, we mapped participant responses to task types identified in our prior studies related to cross-device crowdwork [89, 33, 10, 82, 58, 214]. Specific HIT type labels include *Content Generation* [82, 33, 10, 58], *Image Classification* [89, 10, 33], *Image Transcription* [89, 33], *Information Finding* [89, 58], *Qualification* [10, 58], *Survey* [82, 58] and *Text Classification* [89, 58]. An "Other" label was added to accommodate task scenarios that fail to fit within this label paradigm. The labeling process was conducted by two annotators, and inter-rater reliability was determined to be substantially high for Q15.1 ($\kappa = 0.8$) and for Q17.1 ($\kappa = 0.7$) [226].

In contrast to Q15.1 and Q17.1, Q24.2 is more open-ended such that it allows participants to provide responses that are not necessarily limited to, but may include HIT characteristics. We therefore chose to conduct bottom-up coding process in which themes were developed through standard open-coding. Our underlying intent is to provide an unbiased mechanism for capturing a wealth of characteristics to naturally observe how participants gravitate toward characteristics described in our taxonomy instead of other characteristics that may be relevant (e.g., device constraints). Two annotators engaged in thematic labeling, again, with substantial reliability ($\kappa = 0.8$) [226].

3.3.2 Findings

Demographic information about our participants suggest that they have substantial experience in working on MTurk. Participants identified as male or female near-equally (M=55;F=47;NB=2). 47 participants (45.2%) held at least a Bachelor's degree. In terms of work experience on MTurk, 78 (73.5%) of participants identified as having worked on the platform for 2 or more years. 35 participants (33.7%) stated that they work 10 to 20 hours per week on the platform with a slightly smaller report for the 23 participants (22.1%) who work 30 or more hours per week. The median of total HITs was 26,500 (σ=135,519), and the median approval rating was 99.58% (σ=43.73).

Through our analysis, we observe that that more than half of the survey respondents currently use their phone to complete HITs. Further, we also find that an even larger percentage of participants have explicit ideas for improving mobile HIT usability. 59 participants' responses (57%) made explicit reference to currently completing either at least one of the task types that we described in Section 3.3.1 or HITs that have a particular characteristic that makes them suitable. In contrast to their current practice, 89 participants (86%) provided a response to Q17.1 or Q24.2 that outlined a particular way in which the mobile HIT usability could be improved on smartphones. Across the remaining responses, we observe three high-level themes for using the "magic wand" change how they manage or complete HITs: (1) *Design and Compatibility*, (2) *Task Interaction*, and (3) *Tools, Scripts, and Apps*. Despite being viewed distinctly, each of these themes' responses collectively work toward the goal of improving crowdworkers' efficiency and productivity. The remaining 15 participants (14%) stated explicitly that they do not have any desire to manage or complete HITs on their smartphone. The distribution of responses across these themes is shown in Figure 3.2.

We now present our observations made through the lens of these themes. We draw specific attention to understanding how the characteristics of our taxonomy surface through participant responses. We conclude the presentation of our findings by connecting observations to the types of tasks that are currently practiced by participants alongside those that they believe should be better supported.

Theme 1: Design and Compatibility

As reported by 34 participants (33%), one of the most prominent theme that emerged from our analysis of responses was centered around the resolution of *Design and Compatibility* issues that exist when accessing HITs on smartphones. The general sentiment of these responses was that HITs are designed under the assumption that they be completed on a desktop computer and often fail to render correctly on mobile devices:

> *"Maybe make HITS that just work on that small of a screen. I do other surveys on other platforms on my phone, and they always look better than any of the MTurk one that require a phone." (P38)*

By rendering incorrectly on the phone, the design of the HIT introduces new barriers that require additional effort to complete on the smartphone devices. For example, "Penny HITs" are a common type of HIT that are already divided, are limited in size, and can be completed in less than a few seconds on a desktop computer. Penny HITs often involve snap-judgements about a specific type of *Prompt Modality* with a binary question (e.g., "Is there a cat in this image?"). Even in the case where a task may be divided and design for efficiency, its associated media (e.g., a large image) may appear differently on smartphones, which lead to a hindrance in usability:

> *"I would make it so tasks are easier to see and do on the smartphone. Some of them are not made for smartphone use. So, they end up looking weird and not sized correctly. I would make it easier to do quick penny hits, so you can move through them at a quicker speed." (P79)*

A particular aspect of unresponsive HIT interfaces is that the occlusion of other relevant information on the task interface page is common. Such issues were also mentioned in the case of survey HITs in which P8 described their frustration with the navigational demands that arise through HITs that are not well-designed for use on the smartphone:

> *"[Questions in surveys should have] correct sizing so you don't need to scroll all over with Qualtrics surveys. It's annoying and inefficient."* (P8)

Alongside navigational demands that occur within the limits of a task interface, we find that navigational demands beyond it were voiced as well. Two participants explicitly mentioned challenges related to *Off-Site Knowledge Demands* highlighting the need to "make it easier to switch between browser tabs." (P45) and more generally switch between application windows on the phone:

> *"I worry about going from the MTurk page to the survey page and losing my work. So, I would want to change the process and be able to stay on the same page instead of a separate link to go to.", (P45)*

A small number of participants suggested resolving *Design and Compatibility* issues, such "an auto-reformatting feature that formats HIT pages to better display on tiny phone screens" (P70) or "requiring requesters to design their projects for both [desktop and mobile] platforms" (P6). From the perspective of crowdworkers' the ideal composition of these characteristics for mobile HIT usability is one that makes "it easier to go through questions without stopping" (P25). The vast majority of *Design and Compatibility* issues suggest that mobile efficiency is stunted by tasks that are not well-divided, have context that exceeds the smartphone screen, have prompts that are not supported across devices, and are generally slow to complete in comparison to their desktop counterparts. Each of these confirm the representation of the *Off-Site Knowledge Demands*, *Task Divisibility*, *Task Size Volume*, *Task Completion Time*, and *Prompt Modality* characteristics within our taxonomy.

Theme 2: Task Interaction

Elements of *Task Interaction* were discussed by 20 participants (19%). Responses within this theme centered around interactive challenges that occur within task interfaces on mobile devices. Prior studies have identified a plethora of efficiency challenges for touch-based smartphones [11, 126]. Three participants voiced explicit remarks around touch-based interaction, citing that they would use the magic wand to "just improve interfaces to take better advantage of touch screens" (P63). As P29 states:

> *"I'd make it more realistic to do certain tasks on smartphone, such as a bounding-box tasks, i.e. make it support touch-screen devices."* (P29)

At a high-level, we observe that the limitations in interactivity that stem from touch-based input contribute significantly to our participants' lack of interest in using their smartphone to complete HITs.

Despite not explicitly mentioning the smartphone's touch-based input, the remaining 17 participants provided remarks that highlight how mobile activities are stifled by the speed of mobile interactions. For example, 11 of the 20 participants experienced difficulty when typing with a virtual keyboard on a smartphone:

> *"There is no mouse and keyboard. Everything is slow. It needs to be faster."* (P43)

Seven participants made explicit remarks about the limitations that stem from a lack of traditional input devices. Five of the seven participants further elaborated by explicitly suggesting that the ability to attach a mouse and keyboard to their smartphone would substantially make HITs more usable on the smartphone. As a crowdworker who does not currently complete HITs on their phone, P12 stated:

> *"Perhaps mouse and keyboard support, but this would feel weird. At that point, I guess I may be able to work on a smartphone if forced."* (P12)

Alongside concerns with general efficiency within task interactions, one participant voiced a desire to improve how information is transferred across devices, citing that it was *"hard to copy-and-paste things such as the survey code on a smartphone."* (P5). Across this theme, we specifically observe the presence of concerns and recommendations that span elements of *Interaction Style*, *Prompt Modality*, and *Multi-Device Demands*. Through the lens of *Task Interaction*, our specific observations are that "good" mobile usability is best achieved when task interaction is limited to multiple choice, text entry is not required, and multi-device demands are not present.

Theme 3: Tools, Apps, and Scripts

As reported by 34 participants (33%), the final theme of responses collectively referenced aspects of *Tools, Apps, and Scripts* that relate to the mobile usability of phones. Prior research has found that the vast majority of tools that exist are functionally limited to desktop computers [229]. These tools are often specifically designed to aid crowdworkers in finding and auto-accepting HITs for crowdworkers in order to accelerate their productivity. The entirety of the responses within this theme centered around tools that not only finds and auto-accepts HIT opportunities, but does so in a fashion that targets HITs designed for completion on the smartphone:

> *"I would have an app like I mentioned before that lists only available hits that are smartphone-friendly."* (P64)

Through these responses, we find that crowdworkers believe that there exists a set of HIT characteristics or specific HIT types that make a HIT conducive to complete on smartphones. 21 participants (20%) responded to Q24.2 by explicitly mentioning the need to "mirror the scripts' functionality from my desktop" (P14) in order to find and manage HITs as efficiency on the smartphone. Several participants' responses made reference to the use of the phone was situational and that the smartphone may be used in certain circumstances (e.g., "when I'm not at home" [P63]). In general, these responses not only reinforce the nature of our taxonomy, but also highlight barriers within work practices that exceed beyond the scope of discussion of our taxonomy's characteristics.

Reinforcing Mobile Usability by Task Example

Through our analysis, we observe that crowdworkers hold strong preferences for engaging in specific HITs on their smartphones. Figure 3.3 shows the representation of labels for the HIT types that are currently completed by crowdworkers alongside the HIT types they believe need further support. Reported by 33 participants (31%), we find that Survey HITs are the most prominent type of HITs currently completed on their smartphone, accounting for 47% of the responses to the question. Surveys, in particular, were often accompanied by anecdotal evidence that described their underlying interaction as a motivator for their mobile suitability:

> *"[I'll complete] some surveys that allow it. Anything that does not involve a lot of writing. Some batches that require picking radio buttons."* (P28)

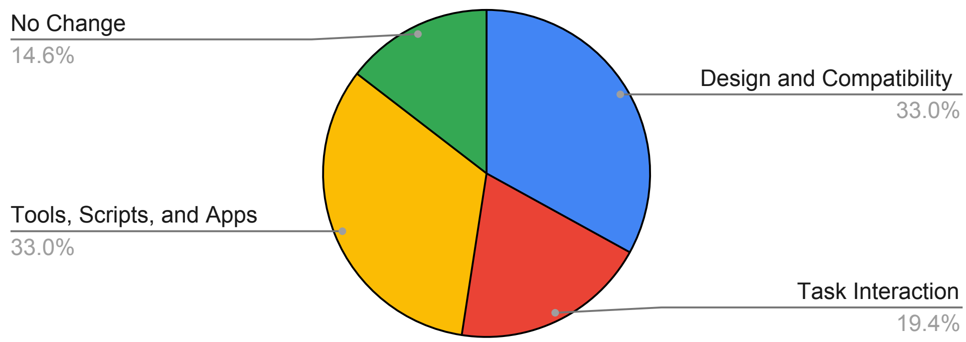


Figure 3.2: Representation of themes in responses to Q24.2.

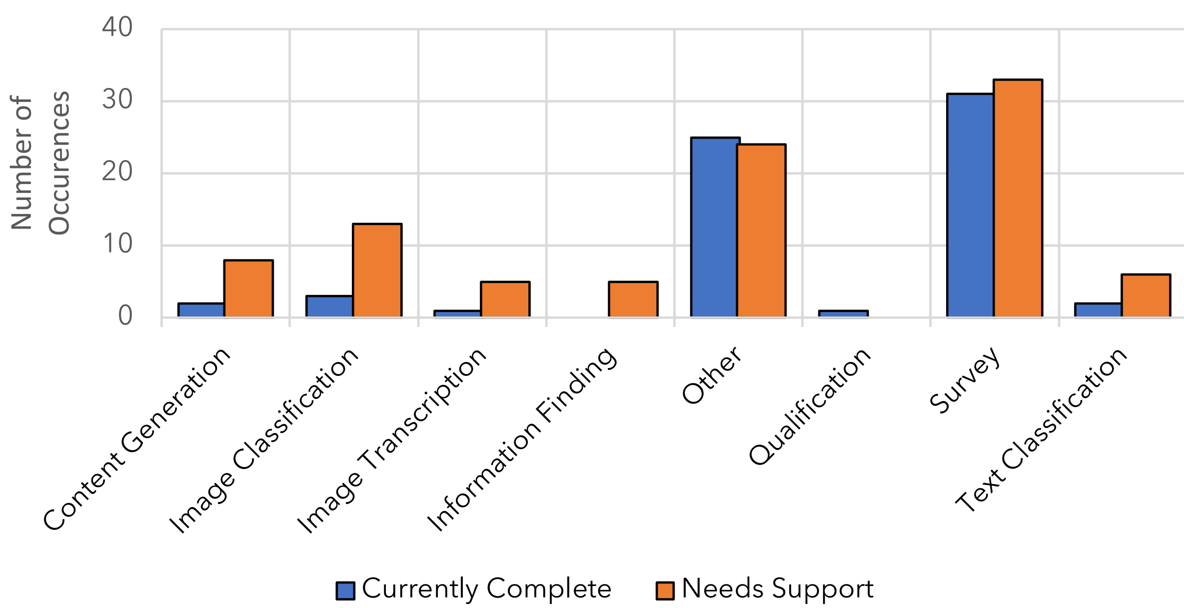


Figure 3.3: Histogram of task types labels that are currently completed or require additional support on smartphones.

Alongside survey HITs, "Other" was the second-most prominent label, reported by 35 participants (33%). Specific responses emphasized the description of characteristics that facilitate mobile usability rather than a specific type of tasks itself:

> *"HITs that look relatively quick. So, [there isn't] much risk if I have to return if they are taking too long. I used to do the dating profile pic HITs as well, but have not recently. I always try to avoid HITs with writing on my smartphone."* (P51)

Other descriptions included "tasks that are simple and don't involve writing" (P87), "quiz-type of HITs" (P60), and "anything with bubble questions to fill out easily" (P25). Several workers referred to these task types as "batch HITs" (e.g., where workers can auto-select from a pool similarly structured tasks from the same requester) and "bubble HITs" (e.g., where workers select from multiple choice answers) Only three participant responses mentioned an explicit type of task, all of which referred to "a lot of app and website testing" (P30).

In contrast to survey HITs and miscellaneous "Other" HITs, the remaining HIT type labels were far less reported. Taken as a collective, our participants engage in tasks that require little navigation beyond the task interface, facilitate quick responses, and require a small amount of time to complete. Mirroring the themes that naturally arose through our coding of Q24.2, we find that much of the reluctance to engage with specific HITs on the smartphone as they are generally not "mobile-optimized":

> *"I will typically only do surveys on a phone if they are mobile-optimized because batch work is impractical on a phone. Most batches use iFrames and want the work done there, which just doesn't translate well to iPhone."* (P29)

Alongside observations related to HITs that crowdworkers currently tend to engage on their smartphone, we observe that crowdworkers not only have an increased desire for currently unsupported HITs, but also have a desire to improve the HITs they're already engaging with while mobile. 34 responses were categorized as voicing a need for further supporting survey HITs while 25 responses were categorized for "Other" types (e.g., HITs that involve downloading files on the smartphone). The specific motivation for these HITs related to the themes that emerged in our analysis of responses to Q24.2, namely *Design and Compatibility*:

"I would like surveys to be better supported on the smartphone. I want them to fit the screen and not involve a whole bunch of scrolling." (P89)

Unlike reports for current counts, participants voiced a need for desiring additional support for all HIT types on smartphones with the exception of qualification HITs, which was not observed in this data in any capacity. Within these specific HIT types, responses often mentioned elements that would improve usability by reducing the effort required to complete a particular task. For example, Image Classification and Bounding Box HITs, which were reported by 13 and 6 participants, could benefit from "easier image bounding" (P33) on the smartphone. Specific recommendations for other HIT types (e.g., Sentiment Analysis) were often not included in responses due to the breadth of the question.

Overall, the findings in this section address RQ1 and RQ2 by illustrating that Mechanical Turk workers do indeed see the same characteristics we have organized from literature.

3.4 Taxonomy Demonstration - Step 3: An Analysis Of HITs In-The-Wild

The results from Step 2's online survey demonstrate that crowdworkers have clear and strong preferences and practices for completing HITs that have characteristics that make them "mobile-optimized". Our qualitative analysis of participant responses specifically suggests that these preferences and practices are molded around the characteristics embodied by our proposed taxonomy. We now aim to conclude our research with a study aimed at using our taxonomy toward assessing the mobile usability of tasks that exist on crowdwork platforms.

3.4.1 Method: Web Scraping and Taxonomy Application

The goal of this study is to demonstrate the utility of our taxonomy. In pursuit of this goal, we build a dataset of HITs using data collected from Amazon Mechanical Turk (MTurk) and leverage our proposed taxonomy as a tool for assessing the mobile usability of the HITs within this dataset. Our approach is specifically inspired by prior studies that manually sample a dataset of HITs from MTurk for manual investigation and analysis [207].

Dataset Generation

We designed web scraping software to scrape HITs and their associated metadata directly from Amazon Mechanical Turk (Figure 3.9). Alongside its collection of basic metadata, the tool was designed to capture two screenshots of each scraped HIT's interface as it renders both in the desktop browser viewport and in the mobile browser viewport. In order to ensure all aspects of the task interface were captured, automated screenshot behavior for both viewports was configured to iteratively screenshot, scroll down, and repeat until the entirety of the vertical visual space had been captured. All information was temporarily stored on two researchers' machines. We developed the tool in NodeJS with Puppeteer[§], a Node library which provides a high-level API to programmatically control a Chrome browser. We collected a total of 519 HITs during June 2021 using a functional worker account on Mechanical Turk.

Qualitative Coding for Taxonomy Characteristics

Using the characteristics in the developed taxonomy as a qualitative codebook, we sought to apply a top-down coding process to the dataset of scraped HITs. The goal of this coding process was to evaluate the usability characteristics of each HIT in our dataset. For each HIT, two coders selected the most appropriate value within each characteristic. When labeling a HIT, for every characteristic in Table **??**, coders selected one of it's values. For example, a HIT's Mobile-Device Demands could be assigned a value of "None" or "Present", based on the definitions in Table **??**.

During this coding process, a subset of 100 HITs (20%) was randomly sampled from the generated dataset of 519 HITs, and two researchers were tasked with labeling each characteristic in the taxonomy. Instructions for labeling involved the use of both the collected HIT metadata and the automatically captured screenshots. Researchers were instructed to make judgements based on what was shown and captured in a HIT's associated screenshots rather than make subjective judgements.

Over the course of labeling the subset of 100 HITs, the two labeling researchers encountered HITs that displayed content that was partially visible or entirely invisible. To account for

[§]https://github.com/puppeteer/puppeteer these scenarios, we developed an eighth label for each HIT in the dataset named “Content Visibility”. Thus, Content Visibility was coded with a bottom-up coding scheme to further our goal of evaluating the mobile usability of HITs in the dataset. This label is relevant to our scraping and coding process and is not included as a characteristic in our taxonomy. Below, we describe the possible values for this label that emerged in reviewing and discussing the sampled HIT instances in the subset of 100 HITs:

- *Visible Content*: The HIT interface is mostly or entirely visible and can be assessed for mobile usability.

- *External Survey Link*: The HIT interface cannot be assessed due to including task instructions alongside an external link to a survey to be completed on a different platform (e.g., Qualtrics, SurveyMonkey, etc).

- *Requester Configuration*: The HIT interface cannot be assessed due to Mechanical Turk algorithmically prohibiting crowdworkers from accessing it via a mobile browser or viewport. An explicit error message is shown in the browser.

- *Acceptance Required*: The HIT interface cannot be assessed due to Mechanical Turk requiring that crowdworkers accept the HIT in order to view the task. Instructions are often shown, but the task itself is not.

- *Data Collection Gap*: The HIT interface cannot be assessed due to missing relevant information stemming from a failure caused by our web scraping software or by the Mechanical Turk platform.

- *Language Mismatch*: The HIT interface cannot be assessed due to being written in a language that is not English.

Following the inclusion of this label, the two researchers revisited the subset of 100 HITs and assigned labels accordingly. For each HIT, researchers first assessed “Content Visibility” as it is a prerequisite for the presence of other labels. If content was visible for a HIT, a label was assigned for each of the seven characteristics. Label agreement was observed to be high across both “Content Visibility” (κ=0.95) and each of the seven labels (κ=0.80; κ=0.92;

κ=0.88; κ=0.86; κ= 0.9; κ=0.97; κ=0.71). Any disagreements or uncertainties were resolved through follow-up discussion. The remaining 419 HITs in the dataset were divided equally among the two researchers to label independently.

After all HITs with Visible Content were labeled with values for each characteristic, we assigned each value label with Good, Fair, or Poor usability. Each value in our taxonomy is associated with Good, Fair, or Poor usability (see Table **??**). For example, the "None" value in the Multi-Device Demands Characteristic is mapped to Good (green). We use this mapping in our findings in charts and to compare usability among HIT types. The taxonomy provides a direct mapping for all characteristic values in our taxonomy except for "Image" and "Text" in the "Prompt Modality". For Prompt Modality, we conditioned Image and Text to have Poor usability if the HIT had excessive scrolling. Otherwise, Image and Text was set to Good usability. This method of mapping usability values (MUR) enabled this research to generate charts that show usability per characteristic and HIT type e.g., in Figures 3.5, 3.6 and 3.8 as well as χ^2 tests.

3.4.2 Findings

Content visibility issues were prominent in the dataset of scraped HITs. Among the dataset of 519 HITs, a total of 261 HITs (50.3%) were labeled as having issues related to content visibility and were therefore impossible to evaluate with respect to usability. Among the HITs which had visibility constraints, HITs labeled as having an *External Survey Link* were among the most prominent, accounting for 136 of the HITs in the dataset (26.2%). The second most common label for problematic instances were HITs labeled as *Data Collection Gap* where the visibility issues occurred related to errors in our web scraping software. The remaining 75 (14.4%) of HITs experienced visibility issues related to configuration issues where requesters disallowed HITs on mobile devices by detecting iOS or Android user-agent strings, requiring the HIT to have been accepted before viewing, or having been written in a non-English language. 20 HITs (3.9%) out of 75 HITs were labeled as a *Language Mismatch* were all written in Spanish. We now present observations from an analysis of the remaining and fully-labeled 261 HITs (50.3%) examining trends in characteristics and HIT types.

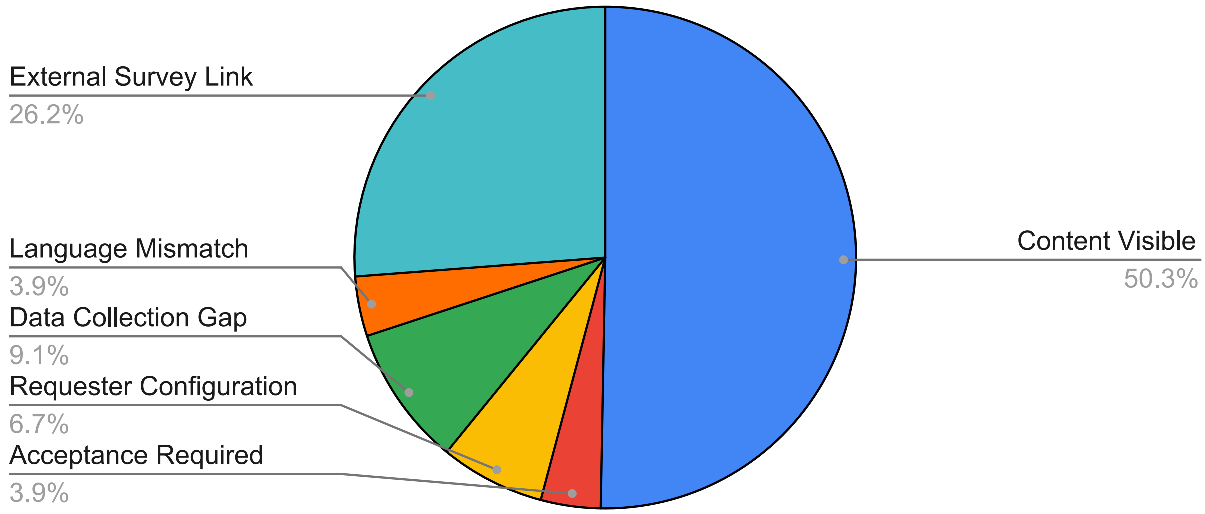


Figure 3.4: Representation of labels for "Content Visibility" across the dataset of 519 HITs.

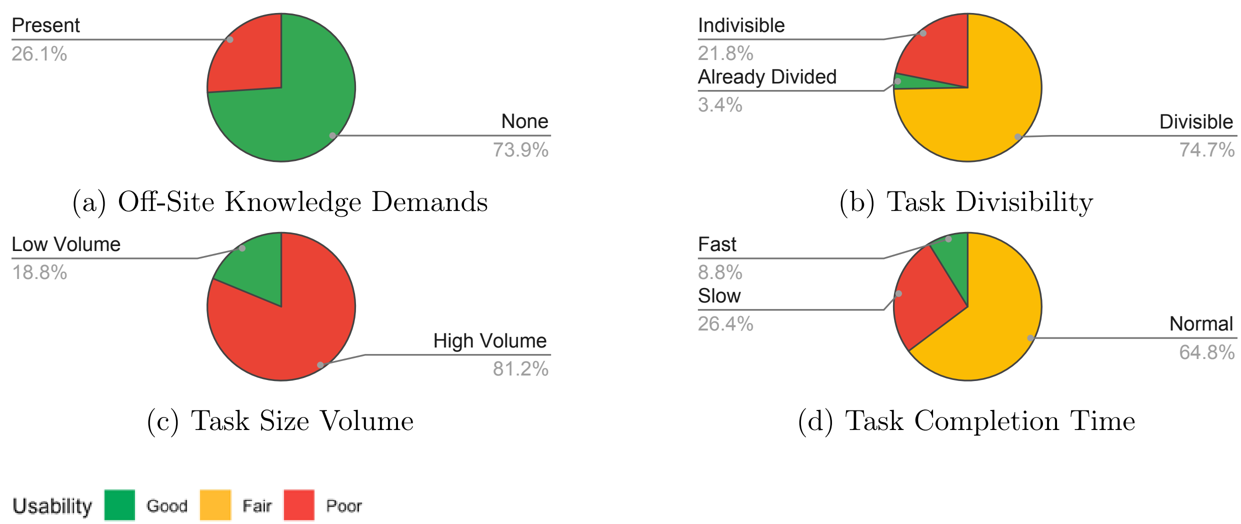


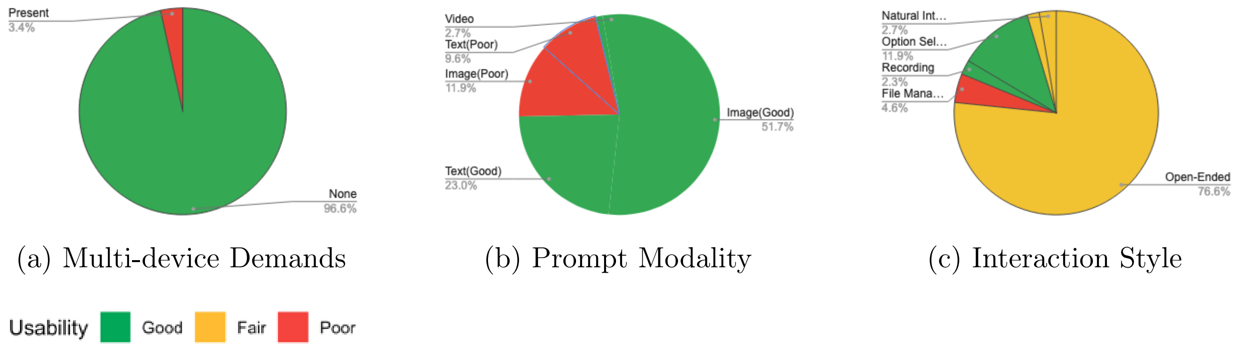
Figure 3.5: Representation of attributes for task characteristics: (a) Off-Site Knowledge Demands with 26.1% Present and 73.9% None, (b) Task Divisibility with 74.7% divisible, 21.8% Indivisible, and 3.4% Already Divided, (c) Task Size Volume with Low Volume with 81.2% High Volume and 18.8% Low Volume, and (d) Task Completion Time with 8.8% fast, 64.8% Normal, and 26.4% Slow.

Assessing Mobile Usability with Task Characteristics

The taxonomy's *Task Characteristics* are *Off-site Knowledge Demands*, *Task Divisibility*, *Task Size Volume* and *Task Completion Time*. Overall, the mobile usability is mixed depending on the particular characteristic. Our analysis suggest that in terms of *Off-site Knowledge Demands* 193 HITs (73.9%) belong to Good Usability as they do not require to navigate away from the main HIT interface. HITs are well suited on smartphones when they have been split into small subtasks. Task Divisibility seems to have the worst usability, as only 3.4% of the HITs are *Already Divided*. In contrast, 21.8% of HITs are *Indivisible* and 74.7% are *Divisible*, but the requesters have not divided them. *Task Size Volume* exhibits even less usability. Only 49 HITS (18.8%) from our dataset have *Low Volume*, leaving 81.2% as High Volume which are less usable on smartphones. Only 8.8% of HITs can be accomplished *Fast*, 64.8% *Normal*, and 26.4% are *Slow*.

Assessing Mobile Usability with Interaction Characteristics

The*Interaction Characteristics* of our taxonomy consist of *Multi-device Demands*, *Prompt Modality* and *Interaction Style*. From scraped HITs, we observe that on average majority of them show *Good Usability* characteristics. It indicate that they would be substantially supported on smartphone. In terms of *Multi-device Demands*, our data show that 252 HITs (96.6%) have *Good Usability* on smartphones as they belong to *None*, while 9 HITs (3.4%) belong to *Poor Usability*. Our data shows that most of the HITs can be highly mobile usable. 205 HITs (78.5%) can be fully supported on smartphone in terms of modality. The remaining 56 HITs (21.5%) that consist of *Image or Text* have *Poor Usability* because of excessive scrolling or occlusion. From our dataset we can see that 247 HITs (94.6%) can be supported using mobile phones with respect to the characteristic *Interaction Style*. These HITs were interacted with using the methods which include *Good-Fair Usability*. They include *Option Selections, Recording, Open-ended+, Object Manipulation* and *Natural Interface Input*. The remaining 12 HITs (4.6%) have *Poor Usability* as their interactions were through *File Management*.



(a) Multi-device Demands (b) Prompt Modality (c) Interaction Style

Figure 3.6: Representation of attributes for interaction characteristics: (a) Multi-device Demands, (b) Prompt Modality, (c) Interaction Style.

Table 3.2: Usability values (MUR) for Good, Fair and Poor and usability percentage aggregated across seven characteristics across each HIT type. We order each HIT type by the Usability Percentage.

HIT Type	Good	Fair	Poor	Usability Percentage %
Image Classification	247	227	93	83.6
Survey	32	13	11	80.4
Text Classification	57	17	24	75.5
Other	63	31	32	74.6
Image Transcription	91	75	58	74.1
Information Finding	200	173	166	69.2
Qualification	6	8	7	66.7
Content Generation	72	32	92	53.1

Assessing Mobile Usability Across HIT Types

Four HIT types – Image Classification, Information Finding, Content Generation, and Image Transcription – accounted for 214 (81.9%) of the 261 HITs. Image Classification is the most the most commonly observed HIT type 81 of the 261 HITs (31%), followed by 77 (29%) *Information Finding* HITs and 32 (12%) *Image Transcription* HITs. The remaining observed HIT labels include 28 *Content Generation* HITs (10%), 14 *Text Classification* HITs (5.4%), 8 *Survey* HITs (10%), and three *Qualification* HITs (10%). A total of 18 HITs, such as text moderation, image quality assessment, and copy editing, are identified as the *Other* category.

Figure 3.8 shows each HIT Type across the 261 fully labeled HITs that represent their distribution of our taxonomy's seven mobile usability characteristics. Figure 3.7 and Table 3.2 show the percent of Good, Fair, and Bad usability aggregated across characteristics. It also includes a Usability Percentage, which we calculate per Task Type by adding the total Good and Fair characteristic ratings by the total number of ratings. This creates a ranking of HIT Types ordered by Good+Fair usability: Image Classification, Survey, Text Classification, Other, Image Transcription, Information Finding, Qualification, and Content Generation. However, if we were to only consider the total Good usability per HIT Type, Text Classification and Survey would have the most usability. This agrees with the HIT types that our survey participants mentioned. Both Text Classification and Survey have *Good usability* relative to other HIT Types, especially their Offsite-Knowledge Demands, Multi-Device Demands, and Interaction Style.

We observe that some characteristics are more commonly identified as "Good" or "Poor" for certain HIT Types in comparison to others. For example, unsurprisingly, Information Finding HITs have Poor usability for the Off-site Knowledge Demands characteristic. Information Finding typically, though not always, involves searching through external web pages. As another example, nearly all Image Transcription and Image Classification tasks have a High Task Volume. At the same time, almost half of Image Transcription and nearly all of Image Classification are *Divisible*, meaning they could be designed in smaller chunks. This represents evidence that requesters could, for example, redesign these two specific HIT types – Image

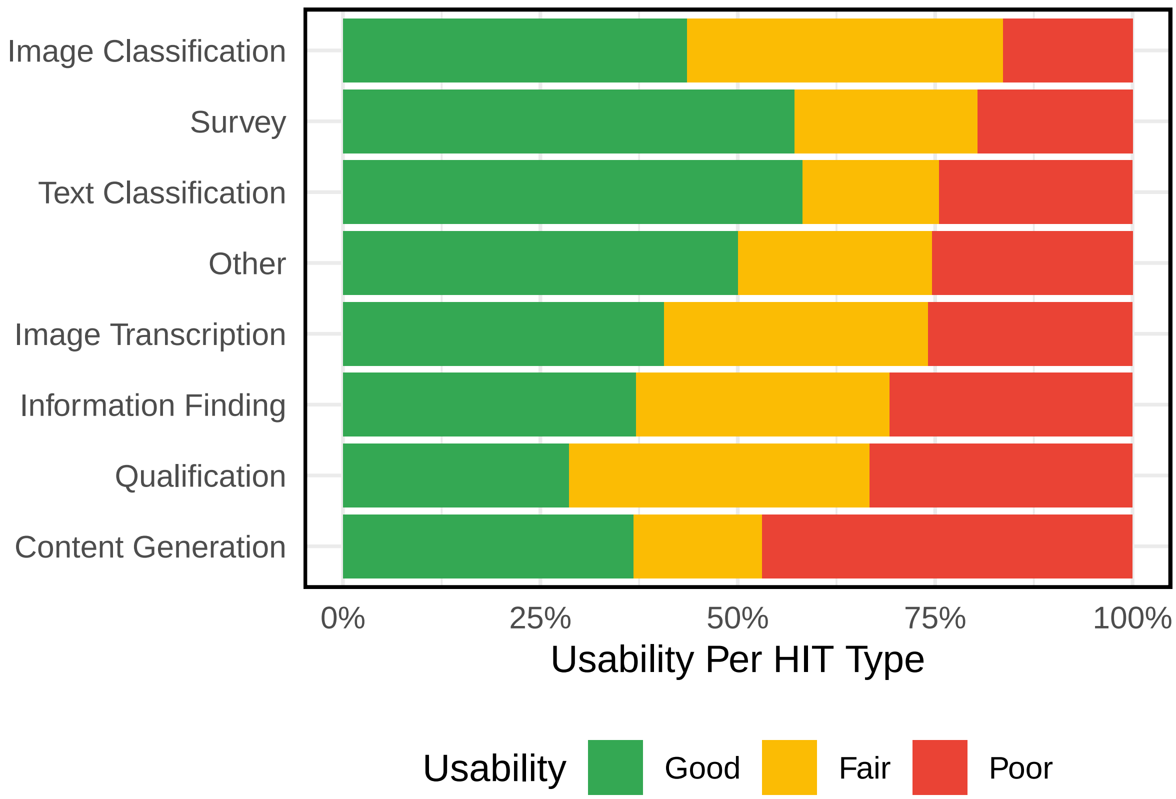


Figure 3.7: Distribution of Good, Fair, and Poor across each HIT type. Here, we order by the least amount of total Poor usability values. This provides a ranking of Tasks, ordered by most to least usable according to our taxonomy.

Transcription and Image Classification – in support of generally improving their mobile usability.

By rank via Figure 3.7, the least usable HIT Types include Information Finding, Qualification, and Content Generation. These show more Poor characteristics relative to other HIT Types, with at least some Poor usability for all characteristics except Multi Device Demands. The remaining HIT Types have mixed usability, having some Good, Fair, and Poor more evenly (though differently) distributed.

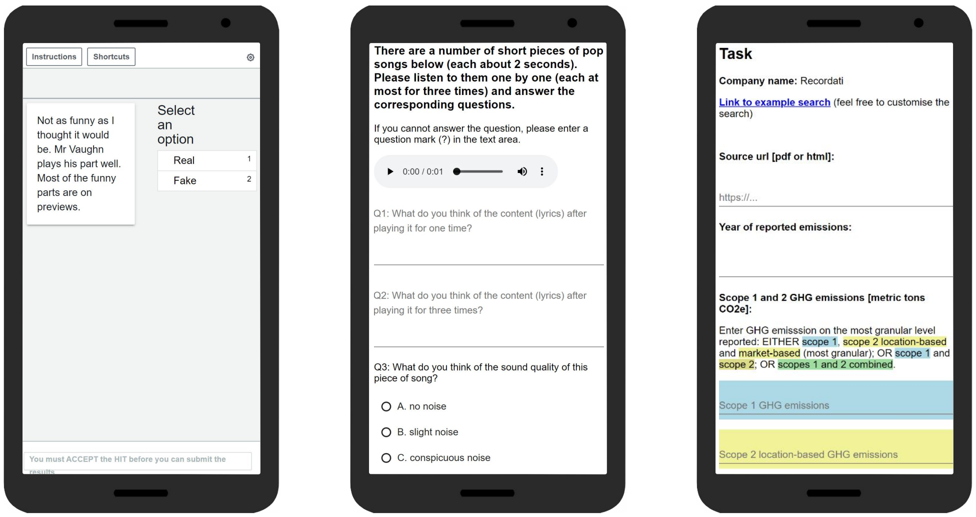
To quantitatively examine how the distribution of usability characteristics vary between HIT types, we conducted a series of Chi-squared tests to test for significant differences. To create task profiles, we binned each unique value-characteristic pair and tallied their occurrences, creating 8 vectors of tallies with 16 dimensions each. We employed a version of the Chi-squared test χ^2 to eradicate issues related to limited observations in our data, that generated multiple simulations to account for smaller data to compare distributions. Each test compared the given distribution against the distribution of the sum of the other (7) profiles. In this case, the null hypothesis is that the given HIT Type's profile is not significantly different from the rest of the HIT corpus' profile. We adjust p-values by applying Bonferroni correction to account for multiple hypothesis testing and report effect size χ^2 (Cramer's V), and the χ^2 statistic. Table 3.3 shows the results of our Chi-squared tests [160].

Overall, six of the eight Chi-squared tests on HIT Types yielded a result that suggests they exhibit usability characteristic distributions that are statistically significant. As shown in Table 3.3, each of the five tests yielded a p-value less than 0.05 with effect sizes ranging from 0.24 to 0.93. Text Classification ($p<.005$), Information Finding ($p<.005$), Image Transcription ($p<.005$), Image Classification ($p<.005$), Survey ($p<.05$), and Content Generation ($p<.005$) all have significantly different usability profiles compared to sum of the other HITs. This suggests that our intuition is correct, that the usability distributions vary based on the type of HIT because of their nature or how requesters tend to design them. These findings, tables, and charts address RQ3 directly, establishing the distribution of characteristics across HIT Types (Figure 3.8) and provides a ranking of hits most conducive to mobile interaction (Figure 3.7).

Figure 3.8: Distribution of Good, Fair, and Poor attributes across all seven taxonomy characteristics binned by HIT type.

Table 3.3: Results from Chi-Squared tests across task types. These tests compare the distribution of each HIT Type's distribution of usability values (MUR) to the other HITs in the set. Significance indicates a relatively distinct profile of usability. (*: $p<0.05$; **: $p<0.01$)

HIT Type	χ^2	β	p	
Content Generation	229.60	0.94	0.004	**
Image Classification	143.26	0.74	0.004	**
Image Transcription	49.41	0.43	0.004	**
Information Finding	163.87	0.79	0.004	**
Other	32.94	0.36	0.095	
Qualification	15.04	0.24	1.000	
Text Classification	79.19	0.55	0.004	**
Survey	44.83	0.41	0.020	*



Task Type	*Text Classification*	*Content Generation*	*Information Finding*
Off-site Knowledge Demands	None	Present	Present
Task Divisibility	Already Divided	Divisible	Indivisible
Task Size Volume	Low Volume	High Volume	High Volume
Task Completion Time	Fast	Normal	Slow
Multi-device Demands	None	None	None
Prompt Modality	Text (No Scrolling)	Audio	Text (Extensive Scrolling)
Interaction Style	Option Selection	Open-ended+	Open-ended+

Figure 3.9: Three example HITs from our scraped corpus with their following Mobile Usability Rating (MUR) based on our taxonomy of characteristics (Figures 3.1 and 3.8). Each has a different Task Type and corresponding MUR ratings. The Sentiment Analysis / Text Classification HIT has an excellent MUR rating. In particular, the Interaction Style and low Task Size Volume is well-aligned for completing on a mobile phone. In contrast, the Information Finding HIT has a poor MUR rating. It requires switching away from the main HIT interface to a website, searching for URLs, and entering the answers with an Interaction Style of Open-ended+. In the middle, the Content Generation HIT can be performed on mobile phone, listening to short clips of pop-music from the same page, but still uses and Interaction Style of Open-ended+.

3.5 Discussion

Our study provides insight on the state of mobile crowdwork. Our work began by developing a taxonomy of characteristics that reflect the usability of HITs on smartphone devices. In demonstrating the taxonomy's utility, we observe that some HIT types – namely Image Classification HITs and Survey HITs – are generally more usable on smartphone devices than other HIT types. Further, we observe that six of the eight examined HIT types exhibit characteristic profiles that are significantly unique from other profiles.

Collectively, our work demonstrates that many HIT opportunities on crowdsourcing platforms, such as Amazon Mechanical Turk, are significantly limited in their mobile usability. These conclusions are further supplemented by trends in preference and practice as self-described by crowdworkers who work on the platform today.

An important consideration for interpreting our results is the characterization of mobile usability. The novelty of the developed taxonomy is grounded in practicality and it ability to surface trade-offs. More precisely, we view the taxonomy as a tool for measuring and tweaking the mobile usability of HITs on mobile devices. For example, in Figure 3.9, the Content Generation task has High Volume, but is divisible. This means the Requester could make it more suitable for mobile smartphone devices by having fewer subtasks per HIT. The same HIT includes an Open-ended+ Interaction Style that can support more nuanced answers at the cost of having less mobile usability. To switch to an Option Selection would make this task faster and more usable, but at the potential cost of data collection needs of the requester.

As shown in Figure 3.4, the taxonomy can be used to assess HITs that are diverse in their nature, spanning a multitude of task types, requirements, and constraints. While each characteristic is designed to represent a particular dimension of a HIT's design, the taxonomy's characteristics are intended to be used in unison. Despite being significantly thorough, our examination of "mobile usability" was conducted to understand usability in the specific context of mobile smartphone devices. We now discuss the implications of the developed taxonomy in the context of mobilizing crowdwork both within and beyond smartphone devices.

3.5.1 Implications for Design: A Taxonomy for Practicing "Good" Mobile Usability

Our research demonstrates how our taxonomy provides a *framework for designing* HITs that are "mobile-optimized". In HCI, Grudin and Poltrock refer to taxonomies as "pretheoretical constructs that characterize cooperative work and identify the technologies that support different types of work" [75]. In our taxonomy none of the characteristics are independently functional, all of them need to be taken together to support mobile suitability. This provides a language for HIT designers potentially strong and weak aspects of designing HITs and balancing both aspects. Requesters (i.e., HIT designers) can utilize our taxonomy and its characteristics as a checklist for ensuring the design of their interfaces are usable on smartphone devices. For example, creating an idealized mobile HIT would avoid *Off-Site Knowledge Demands*, maintain *Already Divided content*, and use *Option Selections* for its interaction style. While some tasks are inevitably more inclined to be less conducive to mobile experiences, our work establishes a design space for future research that seeks to instantiate different HIT designs that leverage the insights of our taxonomy.

Alongside its use as a guideline for mobile design, our taxonomy can also be applied as a *tool for assessing existing HIT design*. Our studies suggest that crowdworkers experience challenges in finding and managing the HITs they accept to be well-matched to their device, setting, and, in some circumstances, their abilities [241]. Uzor et al. suggests that workers that identify as having an impairment would benefit from better market-provider enforced metadata that specifies whether, for example, people with visual impairment would be well-matched to a HIT [214]. There exists a fundamental opportunity for translating our taxonomy of characters into an automated tool as prior studies have previously done in their own usability contexts [3, 220]. Amazon Mechanical Turk and other crowdsourcing platforms could incorporate such a tool to provide requesters with feedback about improving their HIT designs in the same way that interfaces for password creation provide feedback about the "strength" of a password. Further, reviewing platforms, such as Turkopticon [101], could stand to benefit from attaching systematic metadata about the mobile usability of HITs alongside the standard reviewing metadata. The collection of such data could facilitate the creation of data sets of mobile usability ratings, which could be utilized toward automated tools for detecting HITs that have “good” mobile usability.

Alongside general-purpose usability frameworks, there exists an opportunity to explore convergent perspectives of usability and design as we move toward a practice of cross-device crowdwork. For example, Nakatsu et al. [151] developed a taxonomy of crowdsourcing tasks based on task complexity across two dimensions: structure and independence. Our taxonomy presents a thorough examination of mobile HIT usability through well-structured and independent tasks that Nakatsu et al. refer to as “contractual hiring”. There remains a significant opportunity for further characterizing mobile usability in crowdwork that, according to Nakatsu et al., involve multiple parties and significantly less structure.

Finally, our taxonomy can serve as a useful complement to current and future frameworks alike for other niche contexts. For example, significant attention has been given to understanding pathways for improving the accessibility of work opportunities in crowdwork [207]. In a study of accessibility on Amazon Mechanical Turk, Zyskowski et al. found that 39% of respondents reported using an assistive device to engage in crowdwork [241]. Reporting a similar percentage of assistive device usage, Uzor et al. found that these users gravitate toward completing survey HITs in comparison to other HIT types (e.g., information finding) [214]. Though our focus was limited to mobile usability, our taxonomy may be useful to individuals within the accessibility community who have struggled, or continue to struggle, with characterizing accessible HITs. In the same way that “most accessibility fixes actually make products better for all users” [28], we expect that making HITs more usable will positively impact all Mechanical Turk workers.

3.5.2 Toward a Usable Practice of Cross-Device Crowdwork

Our research takes an important step toward defining and evaluating mobile usability for crowdwork. A wealth of prior research at the intersection of crowdwork and mobile computing has repeatedly reinforced the importance and opportunities of smartphone devices [81, 55, 103], which collectively drove us to focus explicitly on understanding usability within this type of device’s context. Our study serves as a foundation for developing and conducting future usability assessments across other types of devices or specific situational contexts.

Recent studies suggest that crowdworkers have an interest in using more devices than their smartphone alone to support their work [87, 229]. For example, Hettiachchi et al. studied task acceptance rates across desktop computers, smartphones, and voice assistants, observing that preferences of task acceptance for smartphones and voice assistants are only slightly smaller than preferences for the desktop [89]. As non-desktop devices continue to become increasingly more important to crowdwork, there exists a growing need to understand how usability should be defined within each of their respective contexts.

By studying, understanding, and assessing device-related usability, crowdsourcing researchers, crowdworkers, and platform can begin to develop new "cross-device" systems, tools, and experiences [156]. *Information work*, for example, encompasses a range of computer-based professions (e.g., programming, design, writing) – many of which are recognized as desktop-centric practices [165]. Among these traditionally desktop-centric professions, studies have shown that people use the smartphone to facilitate communication [95], transmit information across devices [156], and continue tasks more generally across devices (e.g., web browsing [108, 155]). Interactive systems research has focused explicitly on designing new cross-device systems and tools to better understand the benefits that arise from cross-device experiences (e.g., in software development [94] and in every-day experiences [161]). Aligned to our discussion of mobile usability, Mercury [228] and PlayWrite [100] are two mobile microtasking systems that employ microtasks that would, by our own taxonomy's assessment, be deemed as having "Good" mobile usability. Despite being prototypes, each system's evaluation demonstrated its potential for impacting the nature of their respective work context substantially.

We argue that our work provides a framework for defining, understanding, and assessing "cross-device" in the scope of crowdwork: How is it technically feasible for a HIT to be completed across multiple devices? Are there administrative tasks related to crowdwork (e.g., finding HITs) that should also be facilitated across devices? Are there combinations of devices that are more usable with one another than other combinations? Each of these questions poses a particular challenge that collectively hinge on the fundamental characterization of what mobile usability means for a particular device. The landscape of research on cross-device crowdwork remains relatively small, and we therefore encourage researchers, crowdworkers, and marketplace platforms to recognize the area as one that is fruitful for innovation.

3.5.3 The Frontier of Mobile Tooling in Crowdwork

Our study provides insight into the role that tooling can play in facilitating mobile experiences in crowdwork that are both efficient and productive. We specifically find that crowdworkers have a desire to mobilize aspects of their work, but are limited by the tooling that exists today. A wealth of prior research has reinforced the role that workstation-based tools play in enabling efficiency in crowdwork [106, 194, 229]. Kaplan et al. specifically noted that many tools for efficient crowdwork are facilitated through platforms (e.g., Turkopticon [101]) or browser extensions (e.g., MTurk Suite) [106]. Today, Firefox for Android remains the only smartphone-based browser that allows users to load and employ browser extensions. Many of these tools rely on browser APIs that are only supported on desktop browser implementations and are therefore incompatible on mobile devices.

Within the purview of tool development, the role of the crowdworker continues to remain an important consideration. Many of the most long-standing tools in crowdwork, such as Turkopticon [101] and MTurkSuite, are not only worker-developed, but also worker-sustained. For crowdworkers, tools are "the glue that makes their work possible" [229], and this indicates a need to engage with them as researchers that continue to build tools to support them. Further, state-of-the-art workstation tools, such as Otto[¶], require its users to pay a monthly fee of $10.00 per month in order to access the tool's core features. As tool development continues among researchers, crowdworkers, and other participates, there exists a broader challenge of ensuring that tools – whether they be for the desktop, for the smartphone, or any other device – remain available to the public in order to ensure facilitate work experiences that are not only productive, but also fair to crowdworkers at large. This is particularly relevant for crowdworkers (e.g, in rural areas) who may use their mobile device more often than a workstation computer [55, 81, 24, 222].

[¶]https://www.ourhitstop.net/membership-tier-information

3.6 Limitation

This study has several limitations. First, it focuses on mobile crowdwork specifically on Amazon Mechanical Turk (MTurk), limiting the ability to generalize findings to other platforms (e.g., Upwork[‖]). Second, the taxonomy is based on an online survey of 111 MTurk workers, which may be biased by participants' recent activities or memory limitations. To address this, we conducted a follow-up study analyzing 519 HITs from MTurk, though this sample size is still limited. Our findings are context-specific to current mobile devices and HIT types. Future studies can add nuance to these findings by employing more fine-grained methods and tools (e.g., activity logging apps for mobile devices) toward the targeted goal of assessing mobile task preference in practice.

[‖] https://www.upwork.com/

Chapter 4

Understanding crowdworkers' perspective for using various device types in crowdwork

4.1 Introduction

Crowdwork is attractive to workers in part because of the flexibility it provides. Despite the importance of flexibility to workers, the current design of crowdsourcing platforms and the HITs found thereon often limit flexibility, leading to frustration for workers [127, 236]. Recent research has suggested that one avenue for improving workers' flexibility is to make it easier for them to manage and complete HITs using non-workstation devices [44, 45]. While prior research has shown that users are open to the concept of using non-workstation devices for crowdwork [90, 68, 157, 66, 217, 26, 2, 106, 88, 89, 87, 44], it is unclear to what extent and in what manner they are already doing so. Moreover, it is unclear what users expect from these devices or what challenges impede them from more completely integrating these devices into their crowdwork practices.

In this chapter, we seek to fill this knowledge gap by surveying 148 Amazon Mechanical Turk (MTurk) workers about their usage and expectations regarding four non-workstation devices [44]: smartphones, tablets, smart speakers, and smartwatches. In particular, I explore how participants currently use these devices for crowdwork, how they wish they could use these devices, and what challenges impede that desired usage. For each of these topics, we investigate how factors related to (a) location [89], (b) stage of work (HIT management vs. completion) [106, 229], and (c) type of HIT [44, 89] impact device usage and satisfaction.

Contributions from our work include,

1. To our knowledge, our work is the first to quantitatively measure the extent to which smartphones, tablets, smart speakers, and smartwatches are used to manage and complete HITs on MTurk. We find that up to a quarter of our participants are already occasionally completing tasks on these devices, including in rare cases on smart speakers and smartwatches. This confirms the viability of research exploring HIT completion on non-workstation devices.

2. We find that half of our participants are interested in using smartphones and tablets to manage HITs, with one-fifth interested in using smart speakers and smartwatches for this same purpose. We also identify the factors that impede their ability to do so, namely the lack of availability of management tooling on non-workstation devices. If such tools were available, workers would greatly appreciate the flexibility of catching and accepting HITs when away from their desks and homes.

3. We discover problems with the crowdworker ecosystem that are causing significant concern and frustration for crowdworkers. This includes issues regarding the increasing number of bots on the platform, the time provided to complete HITs, and the challenge of finding work that pays a sufficient wage. These are all issues that can be addressed, at least in part, by improving the design of HIT management and completion on crowdwork platforms.

4. Based on our findings, we identify recommendations for improving crowdwork on non-workstation devices. This includes recommendations for HIT completion—targeted at the researchers requesting HITs to be completed—and HIT management—targeted at crowdwork platforms and tool designers.

4.2 Study Design

In our research, we conducted a survey of 148 MTurk workers, exploring how non-workstation devices are used to support HIT completion and management. We also investigate the challenges users face using these devices and how they wish they could use them in an ideal world. The goal of this research is to identify how crowdwork can be improved to better support non-workstation devices and increase workers' flexibility. In this section, we describe the survey's design, the study's execution, and the methods used to analyze the collected data.

4.2.1 Survey Design

Our IRB-approved survey includes 43 questions across five sections which is attached in the Appendix A.

Section 1: Demographics In this section, we collected participants' demographic details, including age, gender, and educational background. We then asked about their experience on MTurk, including their number of completed HITs, hours worked a week, and HIT approval rates. This section concluded with questions about the specific tools they employ for managing and completing HITs.

Section 2: Understanding HIT Completion In this section, we asked participants how often they use a workstation, smartphone, tablet, smart speaker, or smartwatch to complete HITs. We also investigate the impact of the location on HIT completion [89] by having participants indicate how often they use the non-workstation devices to complete HITs when they are at their workstation, at home, or away from home. Next, we asked participants open-ended response questions to state the types of HITs they currently complete using non-workstation devices. This was followed by having them rate (on a scale of 1–5; where 1 represents least usable and 5 represents most usable) the usability of completing traditional AI-training-based HITs identified by prior work [89]: sentiment analysis, information finding, audio tagging, speech transcription, image classification, bounding box on all the studied device types. Furthermore, we asked open-ended response questions about what HIT types they wished were better supported for all non-workstation devices including smartphones, tablets, smart speakers, and smartwatches. We concluded this section by asking crowdworkers another open-ended response question about their motivation to choose which device they will use to complete a HIT.

Section 3: Understanding HIT Management In Section 3, we asked participants to indicate how frequently they used all the mentioned devices for HIT management. They also rated (on a scale of 1–5; where 1 represents least usable and 5 represents most usable) the usability of completing HIT management tasks (identified in prior work [229, 159]) on these devices: finding HITs, auto-accepting HITs, creating catchers/watchers, listening to catchers/watchers, talking to other MTurk workers, talking to requesters. We also asked them to indicate which of these HITs they would like to see better supported on non-workstation devices.

Section 4: Broken Desktop In this section, we inquired about the inconvenience participants would experience if their primary workstation was rendered non-functional. We then asked participants open-ended response questions about which alternative device they would use in this situation and why. Next, participants shared how their workstation and alternative device compared in terms of acceptability and effectiveness. We also asked how easy it was to transfer from the workstation to the alternative device.

Section 5: The Magic Wand In this section, we posed a hypothetical scenario to participants. For each device type (e.g. workstation, smartphone, tablet, smart speaker, and smartwatch), we asked participants open-ended response questions about how they would imagine or optimize any facet of crowdwork related to completing and managing HITs if equipped with a magical wand to do so.

Survey Development

After creating an initial version of our survey, we improved it over several rounds of internal review from our team. Once we produced a version we were satisfied with, we submitted for and received IRB approval for our study. We piloted the survey with a group of 20 individuals. Finding no substantial issues, we proceeded to launch the finalized survey.

4.2.2 Study Execution

After being approved by our institution's IRB, we deployed our survey as a HIT on MTurk on November 9, 2021. Participants were compensated USD $5.00 for completing the study. The average completion time was 20 minutes resulting in an average compensation of $15/hour.

During the qualitative data coding process, we identified two respondents who provided single-word responses to all open-ended response questions, which lacked coherence. Consequently, we excluded these two participants, resulting in a final dataset comprising 148 participants for our analysis. When quoting from participants, we label them as P1–150, based on their position in the data before removing the two invalid responses.

4.2.3 Analysis Methods

The multiple-choice survey questions were analyzed using quantitative methods, specifically through the computation and reporting of response frequencies. For the open-ended response questions, we analyzed the data using a methodology inspired by grounded theory [204]. All coding was completed by two researchers who were present at all stages of the process. One researcher currently holds a PhD in Computer Science while the second researcher is a PhD student in Computer Science. Both researchers have significant expertise and published research papers in HCI. First, these researchers read through each response together, applying open coding to generate an initial set of codes describing the data. If there were ever disagreements about the codes to assign, the coders would discuss them until they agreed. Throughout this process, they used the constant comparative methods [64] to identify codes that were originally separate and could be combined, as well as codes that were originally combined and should be split.

Second, after open coding was finished, the coders conducted axial coding, grouping related concepts, generating themes, and describing how the concepts related to the themes and each other. We do not continue through selective coding (the final step of grounded theory). Through this process, the two coders kept a detailed set of research notes. These notes aided in the process of coding but also included insights and lessons learned as the coders completed the coding process. According to grounded theory, these notes are often just as valuable as the actual codes. Many of our findings are contextualized based on the insights found in the research notes.

In the results section, the reported percentages for the motivations behind completing and managing HITs reflect the responses of all 148 participants, as each one provided feedback on what inspired their use of a specific device type. The percentages reported for completing and managing current different types and processes of HITs are derived from the subset of participants who have mentioned using the device for HIT completion and HIT management, as their insights are most relevant for this context. In contrast, the percentages regarding their desire for completing and managing different types and processes of HITs are from participants who mentioned owning the device type for both HIT completion and management, ensuring the opinions are informed and pertinent. Regarding the "magic wand" question, the reported percentages reflect responses from participants based on their ownership of each specific device type.

4.2.4 Participants

Table 4.1 shows the breakdown of the demographics of our survey, including the demographics overall and based on what devices participants owned. Our participant pool was two-thirds male. Nearly half of the participants were between the ages of 25 and 34 (43%). Most held a bachelor's degree (51%), with only 7% holding an advanced degree. Nearly all participants had been on MTurk for at least a year, with most (49%) between two and five years. The majority spent at least 10 hours a week on MTurk, with 21% using it as a full-time (30+ hours) position. Finally, participants on average had twenty-five thousand completed HITs with a median 100% completion rate. Looking at the devices they owned, nearly all had a workstation (99%, n = 147) and the majority had a smartphone (98%, n = 145), a tablet (84%, n=125), a smart speaker (82%, n =121), or a smartwatch (74%, n = 109). There were no meaningful differences in demographics based on what device the participants owned.

Table 4.1: Demographics for the participants taking the study, less those that were removed for quality reasons

		Overall		Workstation		Smartphone		Tablet		Smart speaker		Smartwatch	
Participants		148	(100%)	147	(99%)	145	(98%)	125	(84%)	121	(82%)	109	(74%)
Sex	Male	96	(65%)	95	(65%)	93	(64%)	80	(64%)	81	(67%)	67	(61%)
	Female	52	(35%)	52	(35%)	52	(36%)	45	(36%)	40	(33%)	42	(39%)
Age	18-24	2	(1%)	2	(1%)	2	(1%)	1	(1%)	1	(1%)	2	(2%)
	25-34	64	(43%)	64	(44%)	64	(44%)	55	(44%)	55	(45%)	48	(44%)
	35-44	42	(28%)	41	(28%)	41	(28%)	35	(28%)	33	(27%)	28	(26%)
	45-54	27	(18%)	27	(18%)	26	(18%)	23	(18%)	22	(18%)	23	(21%)
	55-64	12	(8%)	12	(8%)	11	(8%)	10	(8%)	9	(7%)	7	(6%)
	65+	1	(1%)	1	(1%)	1	(1%)	1	(1%)	1	(1%)	1	(1%)
Education	High School Diploma / GED Equivalent	40	(27%)	39	(27%)	40	(28%)	35	(28%)	32	(26%)	27	(25%)
	Trade School	4	(3%)	4	(3%)	3	(2%)	3	(2%)	3	(2%)	2	(2%)
	Associate's Degree	19	(13%)	19	(13%)	18	(12%)	17	(14%)	17	(14%)	17	(16%)
	Bachelor's Degree	75	(51%)	75	(51%)	74	(51%)	61	(49%)	61	(50%)	57	(52%)
	Advanced Degree, e.g. Master's / PhD / JD	10	(7%)	10	(7%)	10	(7%)	9	(7%)	8	(7%)	6	(6%)
Experience on MTurk	Less than 0.5 years	1	(1%)	1	(1%)	1	(1%)	0	(0%)	1	(1%)	0	(0%)
	Between 0.5–1 years	10	(7%)	10	(7%)	10	(7%)	9	(7%)	9	(7%)	9	(8%)
	Between 1–2 years	20	(14%)	20	(14%)	19	(13%)	18	(14%)	17	(14%)	17	(16%)
	Between 2–5 years	72	(49%)	72	(49%)	71	(49%)	60	(48%)	58	(48%)	52	(48%)
	Longer than 5 years	45	(30%)	44	(30%)	44	(30%)	38	(30%)	36	(30%)	31	(28%)
Weekly Hours on MTurk	1–5 hours per week	9	(6%)	9	(6%)	9	(6%)	8	(6%)	9	(7%)	7	(6%)
	5–10 hours per week	23	(16%)	23	(16%)	23	(16%)	20	(16%)	20	(17%)	16	(15%)
	10–20 hours per week	56	(38%)	56	(38%)	55	(38%)	47	(38%)	46	(38%)	44	(40%)
	20–30 hours per week	29	(20%)	29	(20%)	28	(19%)	24	(19%)	24	(20%)	21	(19%)
	30–40 hours per week	18	(12%)	17	(12%)	17	(12%)	14	(11%)	12	(10%)	12	(11%)
	40+ hours per week	13	(9%)	13	(9%)	13	(9%)	12	(10%)	10	(8%)	9	(8%)
HIT Count (Median)		24170		25000		25000		26000		25000		23000	
HIT Approval (Median)		100		100		100		100		100		100	

4.3 Results

In this section, we present the results of our survey. First, we explore the current usage of workstation and non-workstation devices for HIT completion (§4.3.1) and HIT management (§4.3.2). Next, we discuss how participants wish that non-workstation devices could change to better support their needs (§4.3.3).

4.3.1 Current HIT Completion Practices

We find that nearly every participant ($n = 141$; 97%) primarily uses a workstation to complete HITS (see Figure 4.1). A quarter of participants ($n = 41$; 28%) occasionally use a smartphone to complete HITs, with a tenth ($n = 16$; 10%) doing so using a tablet. While HIT completion is rare for smart speakers and smartwatches, it is not non-existent.

Participants gave a variety of reasons for how they chose what device they would use to complete HITs (see Figure 4.2). When explaining why the prefer to use a workstation, many participants cited its generic ease of use ($n = 67$; 45%), speed for completing HITs ($n = 44$; 30%), the size of its screen ($n = 27$; 18%), or its access to a keyboard and mouse ($n = 27$; 18%). As one participant stated,

> *"A full keyboard is needed in most cases."* (P114)

The primary reason to use a non-workstation device was that the HIT required a specific type of device ($n = 44$; 30%). Looking more closely into the types of HIT participants completed using non-workstation devices (see Figure 4.3), we get confirmation that most participants will only HITs on a smartphone if it is required ($n = 66$; 53%):

> *"Mostly whether [the smartphone is] required or not by the requester. Otherwise, I only really use my phone if the HIT requires audio/video collection or uploads, as this is usually easier on mobile."* (P118)

> *"When I'm required to complete them on a smartphone. Most are testability type HITs for mobile web pages, but for a few, I will download an App for testing, but only if I really trust the source of the App."* (P82)

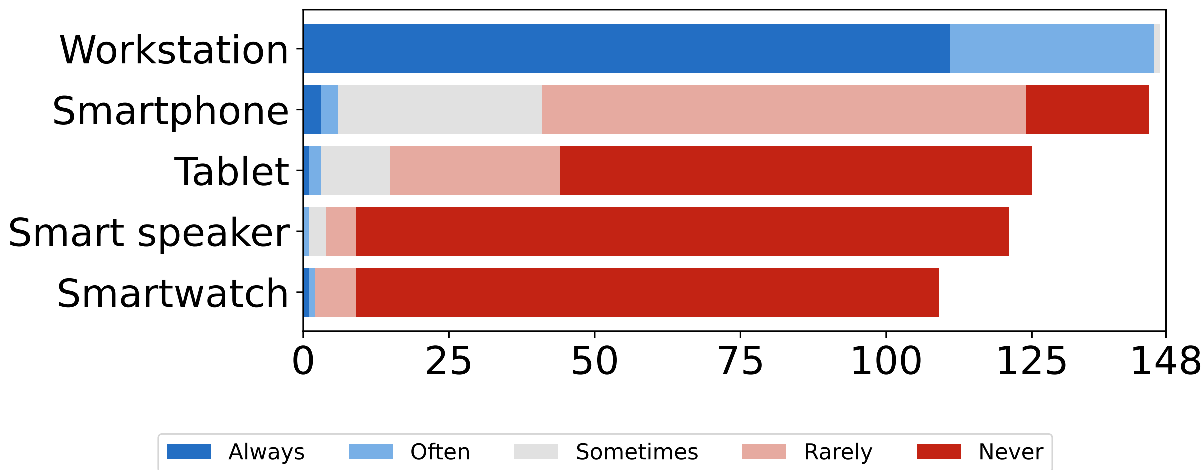


Figure 4.1: Frequency with which different devices are used to complete HITs.

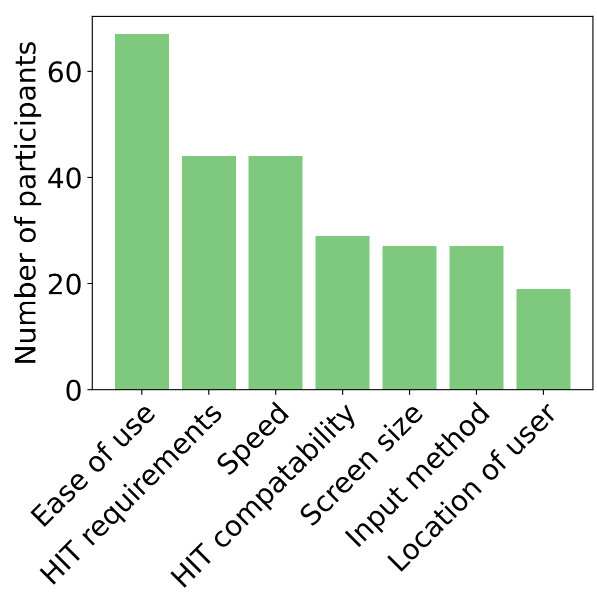


Figure 4.2: Factors Influencing Crowdworkers' Choice of Device for Completing Tasks

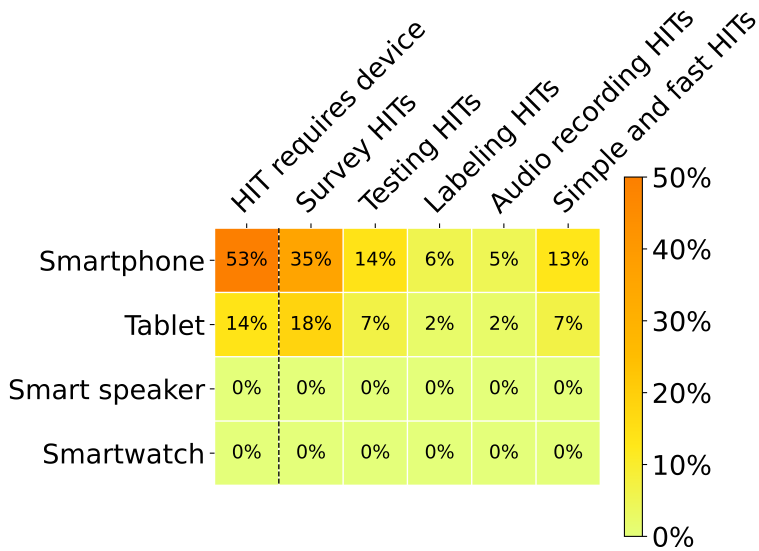


Figure 4.3: Types of HITs crowdworkers are currently completing with different non-workstation devices. Data is from an open-response question. Percentages are based on the number of participants that have used each device to complete HITs.

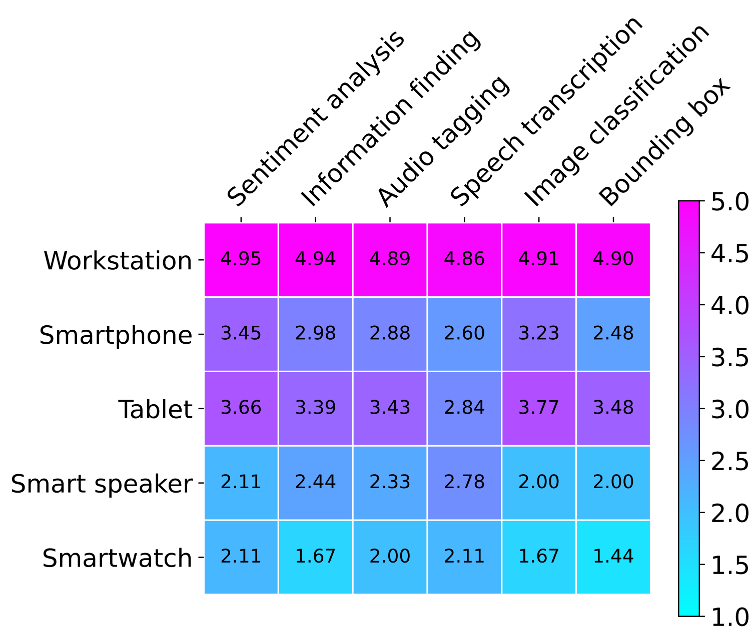


Figure 4.4: Perceived Ease of Task Completion by Crowdworkers Utilizing Various Devices. 5 is high usability and 1 is low usability.

The decision to avoid non-workstation devices makes sense when considering that participants felt it was the most suitable device to complete HITs for all HIT types (see Figure 4.4). Workstations scored at least 4.9 (out of 5) for all HIT types, while the highest score for any other device on any HIT type was 3.7.

To better contrast the performance of workstation and non-workstation devices, we asked participants how they would react if their workstation became unavailable for crowdwork (see Figure 4.5). We found that over half of the participants ($n = 91$; 61%) indicated they would stop working altogether in this situation. This highlights the challenges workers perceive in completing HITs on non-workstation devices.

When pressed to select which non-workstation device they would use if their workstation was broken, participants were split between smartphones ($n = 78$; 53%) and tablets ($n = 70$; 47%).[*] However, participants largely find these alternative devices unacceptable ($n = 95$; 64%), being far less productive than their workstations ($n = 127$; 86%) (see Figure 4.6).

4.3.2 Current HIT Management Practices

Device usage for HIT management is nearly identical to device usage for HIT completion. Nearly every participant ($n = 146$; 98%) primarily uses their workstation for HIT management (see Figure 4.7). A quarter of participants ($n = 36$; 24%) used a smartphone to manage HITs at least sometimes, with a tenth ($n = 16$; 10%) doing so using a tablet. HIT management is rare, but not non-existent for smart speakers and smartwatches.

Participants gave a variety of reasons for how they chose what device they would use to manage HITs (see Figure 4.8). Workstations were preferred for HIT management due to their generic ease of use ($n = 59$; 40%), the tools or features available only on desktop ($n = 38$; 26%), the speed of managing HITs ($n = 36$; 24%), the size of its screen ($n = 15$; 10%), having multiple windows open at a time ($n = 10$; 7%), or having access to a keyboard and mouse ($n = 9$; 6%). Summarizing the benefits of using a workstation, one participant stated,

> *"Once again, efficiency, ease of use, power, and customizability. I feel the desktop blows all of these other devices away in most of these areas. The only other device*

[*] These numbers include participants who stated they would stop working if their workstation was unavailable.

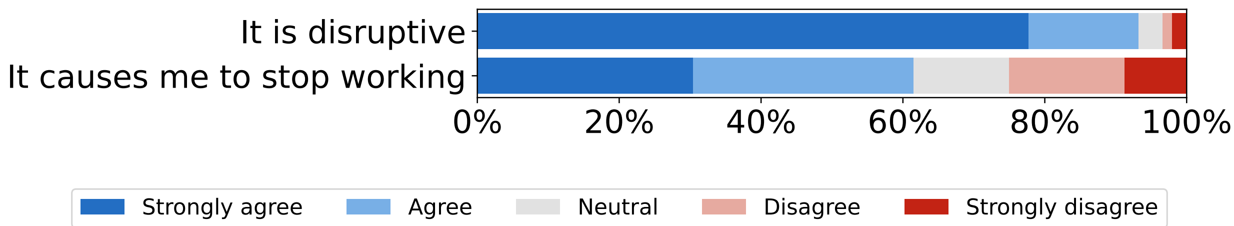


Figure 4.5: Crowdworkers' reaction to the impact of having their workstation break.

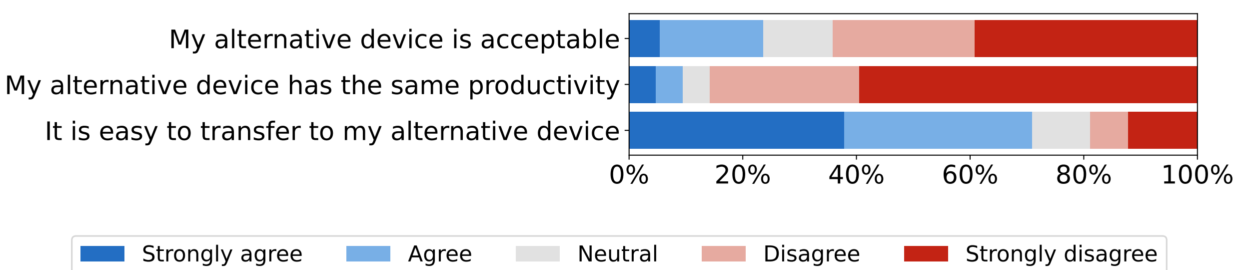


Figure 4.6: Crowdworkers' perceptions on the potential effectiveness of using an alternative device if their workstation breaks.

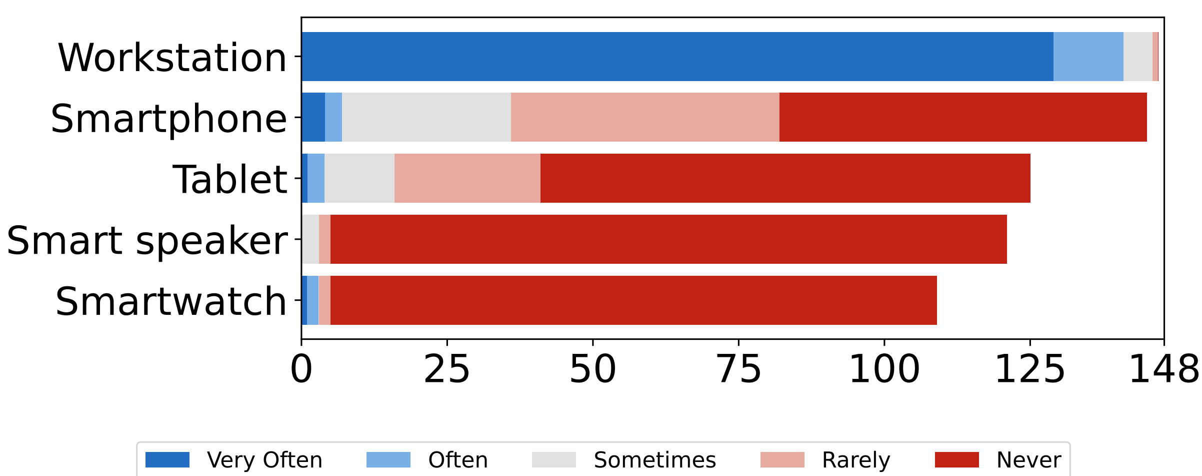


Figure 4.7: Frequency with which different devices are used to manage HITs.

*that comes close would be laptop. However, I would still like to see the *ability* for all of these same features to be available on other devices. I've never tried installing MTurk Suite on my phone... it might work there, but I've never tried it, and I imagine it might not be as efficient as running it on my desktop. It's tough to cram a lot of these complicated features and interfaces on mobile devices."* (P43)

As was the case with HIT completion, the preference for HIT management can at least in part be explained by the fact that HIT management is considered similarly usable on workstations like HIT completion (minimum of 4.8) than on other non-workstation devices (maximum of 3.6) (see Figure 4.9). Looking through all of the participants' qualitative feedback, the coders noted a consistent theme around the lack of availability of HIT management tools on non-workstation devices:

"Only laptop and desktop browsers use scripts or extensions which allow for finding and catching HITs. All the other devices are almost useless in that sense." (P45)

"I use Otto, which is a Chrome extension, and which, as far as I know, hasn't been released on the Google Play store for Android. It finds, catches, and auto-accepts HITs and actually works better than any others I've tried (PandaCrazy, HIT Forker, HIT Finder, etc). I sometimes wish I could get it on Android, as it might be easier for me to just set it and go off to do other things with the Smartphone in my pocket, rather than having to stay close to the laptop to listen for the "beep"!" (P95)

These concerns make even more sense when we consider that the overwhelming majority of participants ($n = 128$; 86%) relied on tooling external to MTurk to manage and complete HITs (see Figure 4.10), with nearly half of participants ($n = 69$; 47%) using at least two tools (see Figure 4.11). The majority of the tools used are focused on HIT management and are not available on non-workstation machines. This helps explain why workers feel non-workstation devices are unsuitable for HIT management.

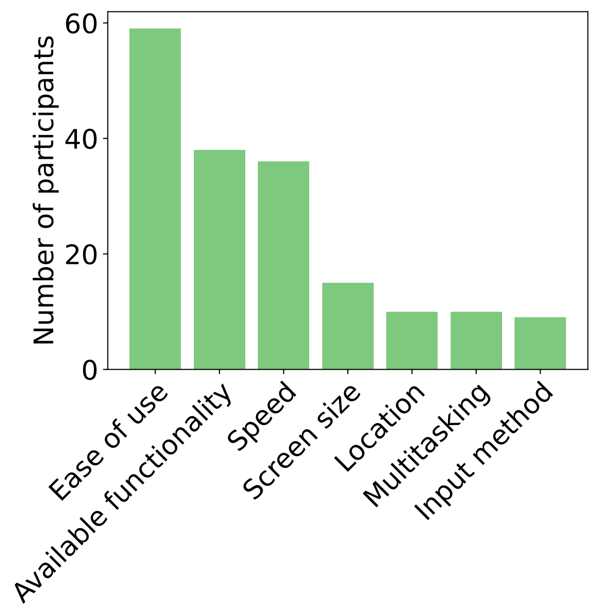


Figure 4.8: Factors Influencing Crowdworkers' Choice of Device for Managing Tasks

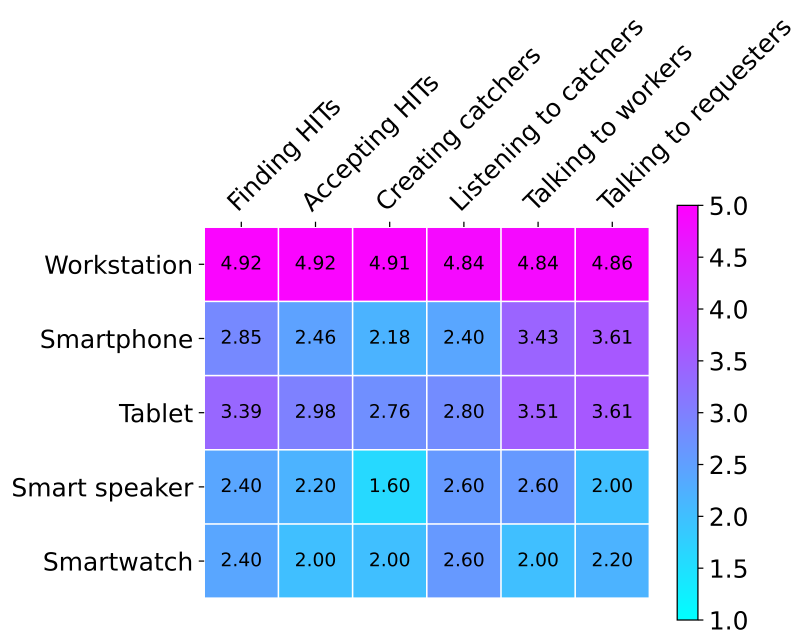


Figure 4.9: "Perceived Ease of Task Management by Crowdworkers Utilizing Various Devices. 5 is high usability and 1 is low usability.

4.3.3 Desired Usage

It is clear that in the current crowdwork ecosystem, participants feel they need to use a workstation to manage and complete HITs. However, whether this situation arises due to satisfaction with workstations or dissatisfaction with non-workstation devices is unclear. To answer that question, we inquired whether participants wanted support for HIT management and completion on non-workstation devices.

For HIT completion (see Figure 4.12 we see marginal interest in using non-workstation devices to complete work, with completing surveys on smartphone ($n = 34$; 23%) being one of the few HIT types with significant interest. In contrast, there is significant demand for better support for HIT management across all non-workstation devices (see Figure 4.13). In this case, roughly half of the participants wanted better support for hit management on smartphones and tablets, while a fifth wanted better support on smart speakers and smartwatches.

To understand what changes would be needed to allow non-workstation devices to be better integrated into their crowdwork practices, we asked developers what they would improve about the non-workstation devices if there were no limitations on what they could request (i.e., they had a magic wand). These requests help illuminate the challenges workers find in using non-workstation devices to complete and manage HITs. Figure 4.14 summarizes their responses.

HIT Management

Half of the participants ($n = 75$; 51%) wanted increased tool availability for non-workstation devices, particularly on smartphones ($n = 54$; 37%) and tablets ($n = 41$; 33%). Many participants viewed these tools as essential for HIT management, noting that the availability of HIT management tooling on non-workstation devices would provide them with significant flexibility:

> *". . . Workers may be away from their workstations or house, but can carry the smartwatch with them, allowing them to catch HITs . . . ."* (P62)

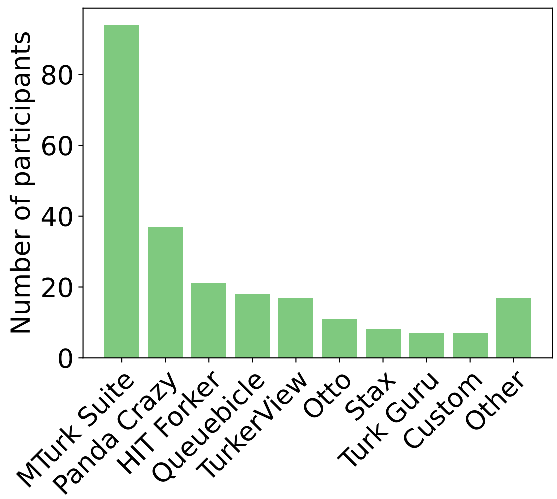


Figure 4.10: Predominant tools utilized by crowdworkers in our study

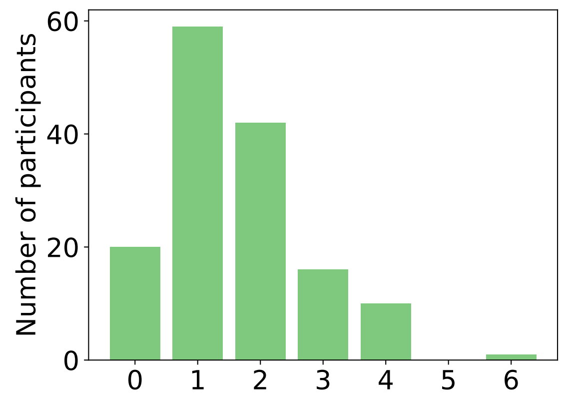


Figure 4.11: The number of tools each crowdworkers uses in our study

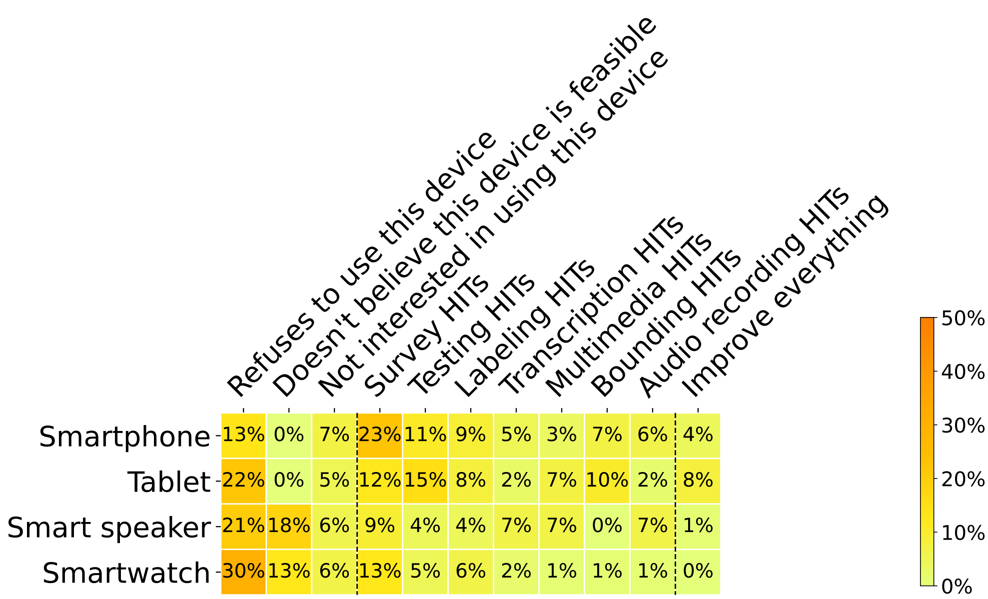


Figure 4.12: Types of HITs crowdworkers from our study are interested in completing with different non-workstation devices. Data is from an open-response question. Percentages are based on the number of participants that have used each device to complete HITs.

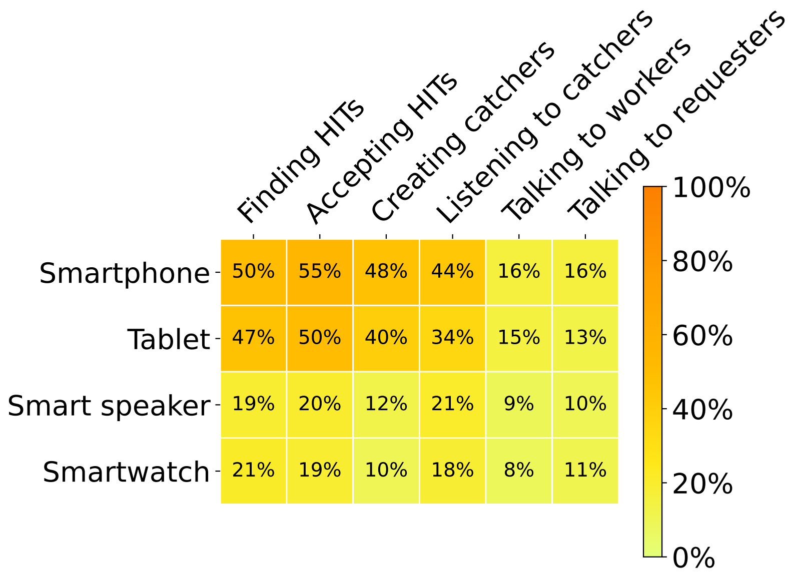


Figure 4.13: The percentage of crowdworkers from our study that want to perform HIT management tasks with different non-workstation devices. Data is from a close-ended question. Percentages are based on the number of participants that own each device.

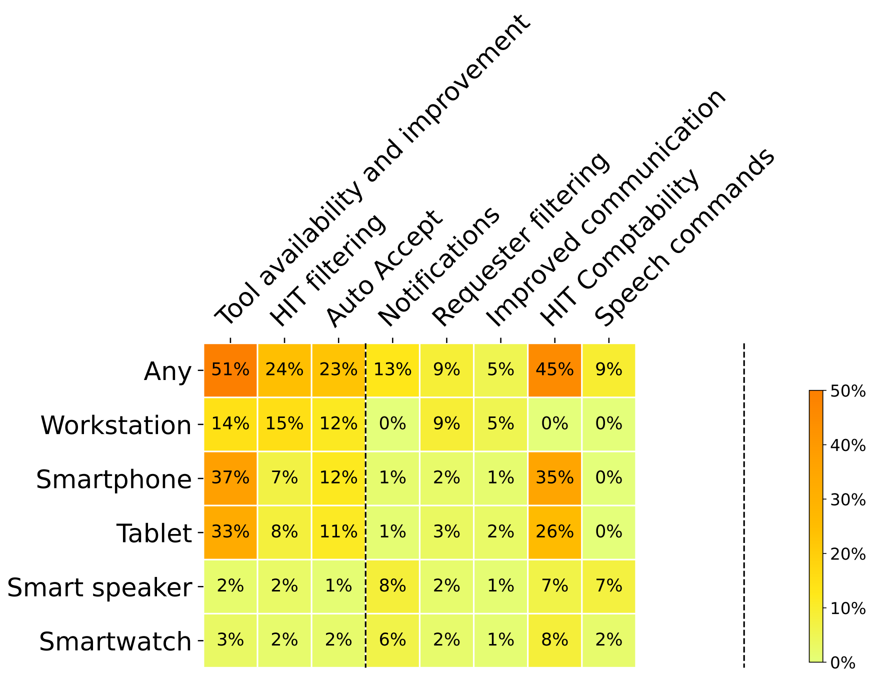


Figure 4.14: The percentage of crowdworkers from our study that indicated various items when asked how they would use a magic wand to improve crowdwork on different devices. Data is from an open-response question. Percentages are based on the number of participants that own each device.

"Setting up alerts and listening/getting a notification for them, being able to speak to accept HITs [on my smart speaker] when I'm away." (P83)

Outside of tooling, participants noted a need for better support for HIT management across all devices. Often, the desire for this tooling was that it would provide better HIT catching ($n = 34$; 23%):

"I would add HIT catchers and scripts to the smartphone platform." (P50)

Participants also want better support for HIT filtering ($n = 36$; 24%):

" I would make it easier to filter out poor paying or low-quality requesters, because I spend a lot of time filtering out junk" (P50)

"Better and easier native ways to categorize and filter HITs and flag or rate requesters." (P119)

Interestingly, while most requests for HIT management improvements are focused on smartphones and tablets, we do see interest in improved notifications on smart speakers ($n = 10$; 8%) and smartwatches ($n = 7$; 6%). For example, participants stated,

"I'd love to be able to hear from my smart speaker anytime a HIT that matches my preferences is available, then ask whether I'd like them to accept the HIT or not." (P12)

"I would allow for notifications from Mturk to go through to your watch to alert you of expiring hits or new hits accepted." (P25)

HIT Completion

Nearly half of the participants ($n = 67$; 45%) wanted requesters to improve the UI and UX of HITs. Such feedback was common for smartphones ($n = 51$; 35%) and tablets ($n = 32$; 26%), and to a lesser extent smart speakers ($n = 8$; 7%) and smartwatches ($n = 9$; 8%). Much of this feedback centered around designing HITs so that they can be more effectively displayed on smaller screens:

> *"I would like [my smartphone] to be easier to accept HITs and to be able to actually do them. I would like for it to be formatted to my phone so I can navigate easier."* (P14)

> *"I would format HITs to fit the screen [on my tablet] . . . This would make it easier to complete tasks and find tasks."* (P87)

Improving HIT design has the potential to improve flexibility by allowing workers to complete HITs when and where they want:

> *"I [would] like to use my smartphone when I'm away from home, waiting for an appointment, etc."* (P40)

> *". . . I will use my tablet if the internet at my house is down because I can use the mobile data on it to complete tasks"* (P95)

Ecosystem Issues

In addition to issues with HIT management and completion, participants also expressed concerns with the crowdwork ecosystem generally.

Bots In our data, we saw that a common reason that participants were worried about HIT management was the perception that MTurk had become overrun by bots, and that only through effective use of tooling could workers manage to catch sufficient work. For example,

> *"I would take the bots away so the real people have a better chance at earning money. It gets really old when HITs are taken in an instant constantly."* (P37)

> *"workers are frustrated by the lack of work provided, especially when considering bots take some of the work."* (P62)

This apprehension greatly influences flexibility, making participants feel anchored to their workstations. It's only at the workstation that they possess the essential tools to compete with bots.

HIT Time Limit Another issue is that participants do not feel as if they have enough time to complete HITs:

> *"I would give workers more time, guarantee individual work based on qualifications. The reasons are that some requesters do not provide enough time to finish HITs, so workers feel rushed."* (P62)

Part of this issue is that participants are using tools to auto-catch HITs. By the time the worker gets to the HIT, the time limit to complete that HIT has already elapsed. This occurs because the clock to complete work starts as soon as the HIT is caught. This can significantly limit flexibility:

> *"Honestly, my only change would probably have to be the timers a requester gives their HIT. if I have to leave my home and end up catching a good HIT while I'm gone, I almost always miss it. I want longer timers."* (P63)

> *"... I lose a lot of work because they're gone within seconds. I'd also magic wand longer timers on HITs ."* (P96)

Wages Lastly, there are also concerns about the wages paid by requesters. Often this was because the time to complete a HIT and the compensation for a HIT were poorly correlated:

> *"I would have HITs rated on $ per hour. I currently auto-catch anything that is above 50c, however, sometimes I see that these HITs are 50c for something like 30 mins, which I find ridiculous. The minimum rate I am usually willing to work is 10c/minute. If I could easily catch any HIT that is over 10c/minute, I think that would help me out a lot."* (P65)

This leads some participants to want more tools that help them learn about the quality of requesters:

> *"I would just improve the existing tools I have, especially Turk Guru. Sometimes it accepts HITs from requesters who have acceptance ratings below my minimum threshold or don't pay my minimum hourly rate. That aspect could be improved." (P68)*

4.4 Discussion

Our study confirms that HIT management and completion are primarily a workstation-only affair [159, 44]. However, our results also show that if non-workstation devices could be improved to better support crowdwork, participants are interested in using these devices, particularly regarding HIT management. In this section, we synthesize the insights and experience gained from conducting our survey to create a set of recommendations for better supporting HIT management and completion on non-workstation devices. We also discuss how our results compare to prior work and the limitations of our study.

4.4.1 Recommendations for HIT Management

Many of our participants ($n = 60$; 41%) spent at least 20 hours a week doing crowdwork, representing a full- or part-time job. Therefore, it is no surprise that they had concerns related to their ability to find sufficient, high-paying work. This challenge is only heightened by the presence of automated bots that workers felt were stealing the best HITs and out-competing the human workers.

To address this challenge, workers rely on scripts and tools that partially automate HIT management. Unfortunately, this ties them to their workstation, as that is the only device where these scripts are available:

> *"I think a desktop is required in the current landscape due to low availability of quality work - using any of the other devices puts a worker at a significant disadvantage so these supporting [HIT management] tasks on a different type of device still don't reach desktop quality."* (P71)

An obvious solution to this problem would be for tool makers to figure out ways to bring support for their tools to various non-workstation devices. *We strongly recommend this course of action, as our results show that it is clearly desired by crowd workers and would bring them immediate value.* Similarly, the crowdwork platforms—MTurk in this case—could implement HIT management tooling directly into their website, making it easily available across a wide range of devices.

Another common concern is that it is difficult for workers to easily determine which work can provide a living or reasonable wage. To address this problem, crowdwork platforms could allow workers to filter and sort HITs based on the requesters' acceptance rate and average hourly pay of the requesters. Requesters can already filter workers based on similar criteria, so it only makes sense that workers should also be able to filter requesters. Similarly, crowdwork platforms could require minimum payment rates (per hour), and penalize requesters who repeatedly violate the minimum—for example, forcing the requester to make an additional payment to workers or banning them from the platform. We acknowledge that Prolific has already taken steps in this direction,[†] and encourage other platforms to follow their lead.

Below, we provide several recommendations for how specific non-workstation devices could be used to support HIT management. As all of these devices have limited (or no) screen space, care will need to be taken to design tooling to employ a minimal, clean interface, a simple navigation structure, and multiple modes of interactions where possible [6, 203].

Smartphone While setting up HIT catchers might be better suited to a workstation or tablet environment, smartphones could excel at notifying workers when they have caught work, allowing them to accept that work regardless of where they are at the moment. Similarly, as users are already accustomed to using their smartphones for communication tasks, they would likely also be an ideal avenue for supporting communication between workers and requesters. It might even be possible to integrate with popular messaging apps (e.g., iMessage, Discord) to provide seamless communication. Finally, as smartphones have sensors not necessarily available on workstations, it would be helpful if crowdwork platforms provided a method for easily filtering HITS based on the sensors they require, allowing workers to find these types of specialized HITs when working from their smartphones.

Tablet Compared to other non-workstation devices, tablets have larger screens, higher performance, and the potential to dock with a keyboard and trackpad/mouse. As such, in many cases, it should be relatively easy to directly port existing HIT management tooling to the tablet. Doing so would immediately improve workers' flexibility, as a tablet is much

[†]`https://researcher-help.prolific.com/hc/en-gb/articles/360019777180-How-do-I-resolve-underpaying-st`

easier to carry around than a workstation (or laptop). However, tablets also have many if not all of the sensors found on a smartphone, so they would also benefit from tools that help them filter and find HITs that rely on these sensors.

Smart speaker Smart speakers could be used to notify workers when they have caught work, giving participants the confidence to step away from their desks. Additionally, smart speakers could be used to set reminders related to crowdwork or schedule accepted HITs. When designing crowdwork-specific voice commands, tools should prioritize short commands rather than lengthy dialogues [132]. We think that there is significant space for future research investigating what other ways smart speakers might be able to support HIT management.

Smartwatch Smartwatches could excel at promptly notifying workers of caught work. However, due to the small screen size found on smartwatches, future research will be needed to identify what information to prioritize when displaying notifications. Similarly, care will need to be taken to identify mechanisms for workers to easily react to their notifications, such as accepting or rejecting caught work. Tooling on smartwatches could also consider how haptic feedback, voice input, and hands-free interactions could be used to support displaying and reacting to notifications [16, 181, 114]. As with smart speakers, we think that research into smartwatch-based HIT management is rife with potential.

4.4.2 Recommendations for HIT Completion

A non-negligible amount of our workers already complete HITs using their smartphone ($n = 141$; 97%) or tablet ($n = 16$; 10%) to complete HITs. However, these same participants identified challenges they faced completing work using non-workstation devices. Most commonly, HITS either explicitly prohibit the use of certain devices or are designed in such a way that they are functionality unusable on non-workstation devices.

To address this challenge, we recommend that HIT requesters (i.e., researchers) leverage the principles of responsive web design. That is, wherever possible, HITs should be designed to automatically modify their layout and UI based on the size of the screen where the HIT is displayed. To support these efforts, we recommend the usage of existing responsive web frameworks such as Bootstrap[‡] or Pico.css.[§] Helpfully, modern browsers allow for the testing of web interfaces using a variety of simulated devices, allowing HIT requesters to confirm that their responsive designs work as intended.[¶] Moreover, there are already copious guides on using these frameworks and support is available through platforms such as StackOverflow.

Another common concern is that workers do not have sufficient time to finish the HITs they have accepted. This arises from the use of auto-catchers, leading to workers accepting work as fast as possible and then completing it at their leisure. While in some cases, researchers may need immediate responses, in most cases we feel this is unlikely. As such, we recommend that requesters consider significantly lengthening the time workers are given to start HITs after they are accepted. Based on participant responses, we hypothesize this might also lead to more responses from human respondents (i.e., not bots), increasing the ecological validity of research results.

Finally, we think there is substantial room for more contextually aware HITS [37, 69] that take into account the devices workers are using, the location they are completing HITs, and the social situations they are in [89]. With this in mind, we detail below how HITs could be designed to better fit non-workstation devices.

Smartphone Smartphones have an array of sensors—e.g., accelerometer, gyroscope, magnetometer, GPS, biometric sensors, camera, and microphone—rarely found in workstations. HITs can be designed to specifically target smartphones to gain access to these sensors. For example, Feng et al. built a CAPTCHA system based on the accelerometer and gyroscope, validating it through a user study on MTurk [51]. We think there is room for many more studies along this vein. Potential examples could include,

- Creating a long-running HIT that uses participants' GPS, accelerometer, and/or gyroscope to measure data about how users move throughout the day.
- Using the microphone to measure noise levels in an environment and measure how that impacts task completion.

[‡] `https://getbootstrap.com/`
[§] `https://picocss.com/`
[¶] https://developer.chrome.com/docs/devtools/device-mode

- Measuring how users interact with websites when using a touchscreen.

- Using the GPS to confirm participants' eligibility to participate in a study.

Of all the HIT types that can be completed with non-workstation devices, completing surveys on smartphones was the only HIT type that had significant interest ($n = 34$; 23%). However, participants also note that it is annoying when they encounter open-response questions on their smartphones. We think this could be addressed by leveraging the microphone to record users' responses when they are on a smartphone and using speech-to-text functionally to transcribe their answers.

Tablet As tablets most closely resemble workstations, they are already well-positioned to be used for HIT completion. However, as with smartphones, there is room to leverage the unique sensors found on smartphones. As laptops and smartphones are so similar in their functionality, differing primarily in screen size, researchers could create HITs that compare the performance of tasks on smartphones and tablets, using this to isolate the impact that screen size has on the task.

Smart speaker The key to designing for smart speakers lies in leveraging the high-quality microphones and speech-to-text functionality found in these devices. For example, we believe that audio transcription could be effectively carried out on these devices, similar to what Vashistha et al. did with smartphones [217, 218]. Also, when designing HITS for smart speakers, it will be necessary to provide users with direct and succinct audio prompts that guide them through completing the HIT [61]. We do not recommend designing HITS that will require the smart speaker to describe a web page, as screen readers are notoriously problematic [124]. Compared to any of the other non-workstation devices, the most work will likely be necessary to create HITs that work well on smart speakers.

Smartwatch While for some studies, responsive web design will be sufficient for the HIT to be completed on a smartwatch, in other cases, more tailoring will be needed. Due to their small screens and limited input space, HITs targeting smartwatches should consider incorporating tap-based actions [196], gesture-based input [192, 110, 196], or speech-to-text transcription [125]. Additionally, HITs should consider incorporating the unique sensors found on smartwatches, particularly in regard to health monitoring. For example, HITs could use these sensors to conduct population-wide measures of health data, identify how current health status impacts task completion, or track how the health status of workers changes over time. Any of these could lead to fascinating research in the future. However, significant caution must also be taken to consider user privacy when health data is involved.

As with all non-workstation devices, care should be taken to optimize HITs to run efficiently with minimal battery drain. However, this is likely to be even more of an issue on smartwatches. As such, we recommend that HIT requesters prioritize processing data on the server, not the watch. At the same time, care should be taken to avoid transmitting too much data over the network, which may also impact battery life.

4.5 Limitation

This study has several key limitations. First, our participant pool is limited to 148, which affects the generalizability of our findings. Follow-up quantitative research is needed to determine if these results apply to the broader population. Second, our research focuses solely on Amazon Mechanical Turk, so future studies should explore other platforms like Prolific, CrowdFlower, or Upwork to see if the results are consistent. Third, our results are based on self-reported data, which may include inaccuracies due to misremembering or acceptance bias. Future research should develop prototypes and study actual usage to validate desired practices. Despite these limitations, self-reported desired device usage is important because it provides essential baseline statistics, captures users' subjective experiences and preferences, and identifies gaps between current system performance and user expectations.

Chapter 5

Understanding systematic differences between crowdworkers based on their work practices

5.1 Introduction

One of the crucial reasons crowdworkers like doing crowdwork is the flexibility it offers. But even though flexibility is a big deal for workers, current architecture of crowdwork platforms significantly constrains the flexibility and autonomy that workers possess, especially in orchestrating their work schedules and selecting their preferred work environments [159, 44]. This rigidity starkly contrasts with the inherently fluid and varied nature of crowdwork, underscoring an immediate necessity for these platforms to evolve. Therefore there is an urgent need for crowdsourcing platforms to evolve to better support the varied and dynamic work practices of crowdworkers.

To address this need, it's essential to dig into the core characteristics of work processes that vary among crowdworkers. A standard workflow for crowdworkers includes stages such as managing tasks, which involves finding and accepting work, as well as completing various types of tasks like information retrieval and classification [229, 159, 58]. Previous research has identified device type [45] and work stage, including task completion and management [229], as critical aspects influencing work practices. While prior studies have illuminated crowdworkers' preferences for utilizing a range of devices beyond traditional computers for both completing and managing tasks [89, 229], a significant gap remains in understanding whether there are systematic preferences among crowdworkers for choosing specific device types at different work stages.

In this paper, we aim to bridge this knowledge gap by conducting a survey of 150 Amazon Mechanical Turk (MTurk) workers. We investigate the current practices and desired preferences of crowdworkers regarding the use of various devices. Building on previous research, we focus on four key types of non-workstation devices, in addition to workstations, that have garnered significant interest: smartphones [44, 217, 26], tablets (including situated tablets)[90, 68, 66], smart speakers[88, 87], and smartwatches [157, 2], for both completing and managing tasks. Through this exploration, we uncover three distinct clusters of workers providing a deeper understanding of the diverse ecosystem of digital labor.

Our contributions include identifying three distinct groups of crowdworkers with varying work practices based on device usage and work stages:

- **Cluster 1 (n = 66)**: The largest group, consisting of workers who do not desire to use non-workstation devices for task completion and are highly skeptical about supporting management on all non-workstation devices. This indicates their dependence on traditional devices, though they exhibit limited openness to using smartphones and tablets for completing and managing tasks.

- **Cluster 2 (n = 47)**: The second-largest group do not favor using non-workstation devices for task completion in the future, indicating a reliance on traditional devices. Desite this, they have openness using non-workstation devices currently as well as they want support for scripts and tools to manage tasks across all devices, especially for smartphones and tablets.

- **Cluster 3 (n = 35)**: The smallest group consists of non-workstation enthusiasts who prefer using non-workstation devices, particularly smartphones and tablets, for both managing and completing tasks. Additionally, they express a strong desire for enhanced management support across all devices.

Through these insights, we offer design recommendations for creating more personalized, flexible, and efficient crowdsourcing platforms, fostering a more adaptable and worker-friendly environment. Additionally, we compare the unique work practices of these clusters of crowdworkers with more traditional and established worker groups to better understand their work practices.

5.2 Methodology

In our research, we conducted a survey of 150 MTurk workers, to understand the difference between crowdworkers based on their work practices. We used the same survey mentioned in Chapter 4. In this section we discuss the methods used to analyze the collected data.

5.2.1 Analysis Method

In our previous work we showed all the responses based on HIT Completion and management in Chapter 4. We transformed categorical responses from multiple-choice questions into discrete numerical values. For analyzing open-ended questions, we adopted a grounded theory approach [204]. This analysis was conducted by two researchers, one with a Ph.D. and another a Ph.D. student in Computer Science, both with extensive experience in HCI research. For the organization of qualitative data, we assigned distinct codes to separate columns, utilizing binary values to denote the presence of each code.

PCA was employed to reduce the dimensionality of the dataset, aiming to preserve its core structure and patterns. This step was specifically targeted at analyzing responses related to crowdworkers' current and desired practices for managing and completing HITs, to highlight underlying relationships and diminish data complexity. We normalized both quantitative and qualitative datasets to prepare for PCA. The selection of PCA components as three was determined by using Kaiser rule, focusing on components with eigenvalues exceeding the average to determine their significance as shown in Figure 5.1. Then we select the feature of each PCA component in order to understand what each PCA component represents. We chose the features of each PCA component based on a threshold of the absolute value 0.10 for each PCA component. Choosing a threshold of 0.10 for PCA loading scores ensures that

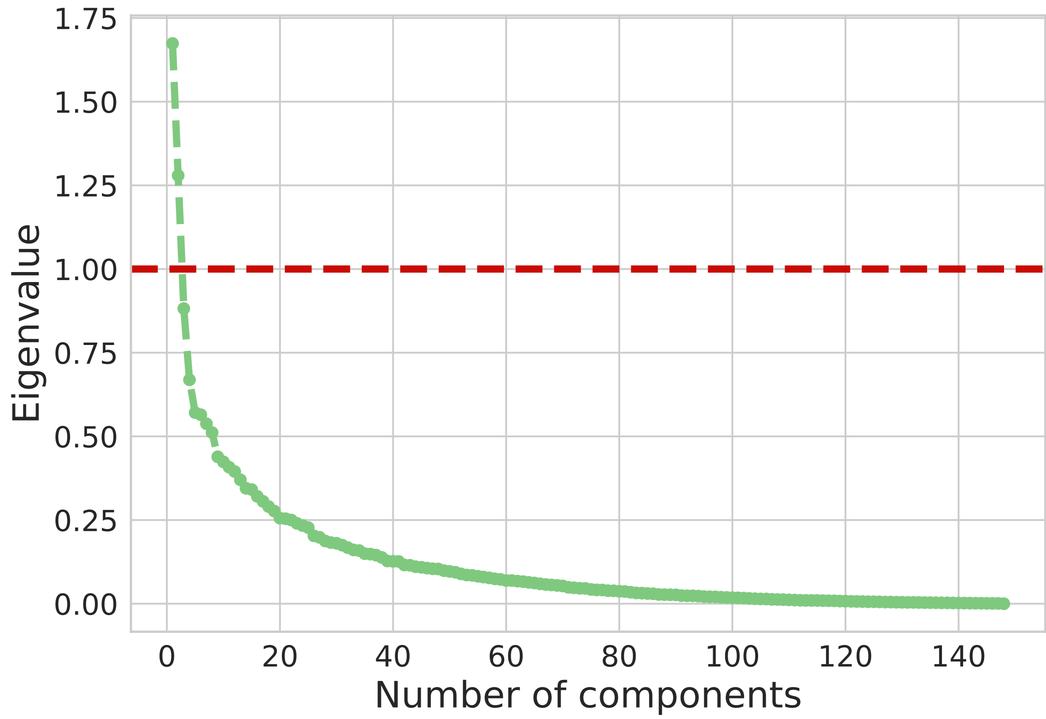


Figure 5.1: Scree Plot which shows the optimized number of PCA components 2 for our dataset

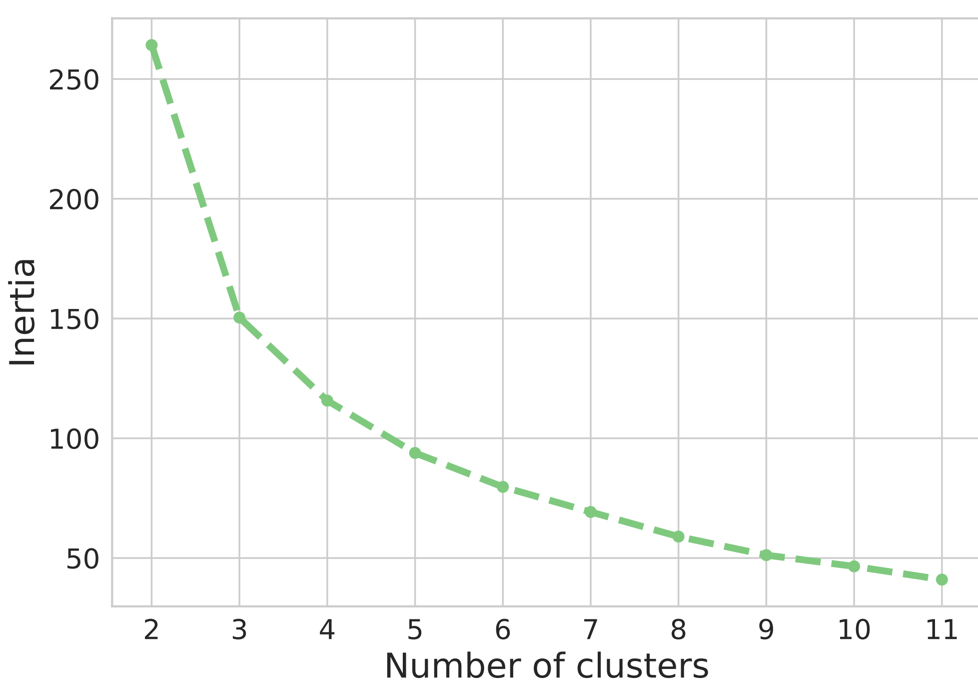


Figure 5.2: Elbow plot which shows the optimized number of clusters are 3 for our dataset

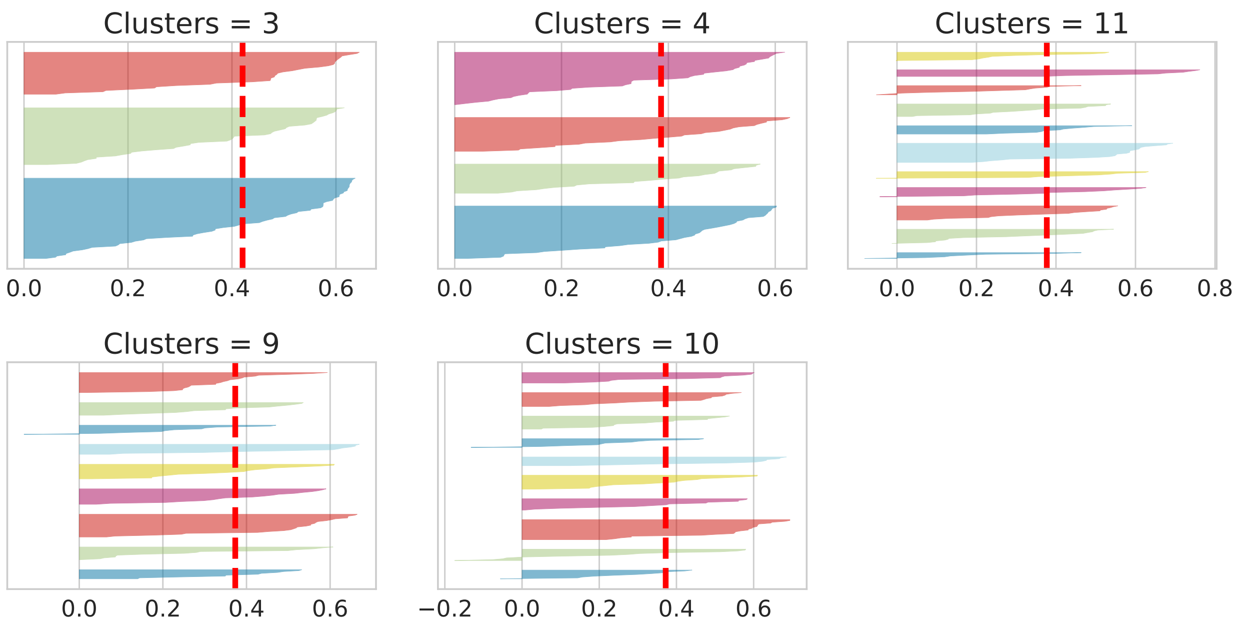


Figure 5.3: Siloutte Score Plot which shows the optimized number of clustering are 3 in our case

Table 5.1: Top features of PCA Component 1 along with their loading factor within threshold absolute value 0.10

Features	**Loading Factor**
BetterSupportTabletManagement	0.21
BetterSupportPhoneManagement	0.20
BetterSupportWatchManagement	0.19
BetterSupportSpeakerManagement	0.19
Magic_wand_Tablet_management_script	0.18
Magic_wand_any_HIT_management_script_on_device	0.18
Desired_HIT_type_Tablet_wont_use	-0.15
Desired_HIT_type_Watch_wont_use	-0.15
Desired_HIT_type_Speaker_wont_use	-0.15
Desired_HIT_type_Phone_wont_use	-0.13
Magic_wand_Phone_HIT_management_script_on_device	0.13
Current_HIT_type_Tablet_dont_use	-0.12
Magic_wand_any_HIT_ux	0.11
Desired_HIT_type_Phone_survey	0.10
Desired_HIT_type_Tablet_survey	0.10
Magic_wand_any_HIT_management_filtering	0.10

Table 5.2: Top Features of of PCA Component 2 along with their loading factor within theshold absolute value 0.10

Features	**Loading Factor**
RateTabletManagement	0.20
RatePhoneManagement	0.18
Magic_wand_Phone_HIT_management_script_on_device	-0.18
RateTabletCompletion	0.17
Current_HIT_type_Watch_dont_use	0.15
Current_HIT_type_Tablet_dont_use	0.14
RatePhoneCompletion	0.13
Current_HIT_type_Phone_device_specific_only	-0.13
Magic_wand_any_HIT_management_script_on_device	-0.12
Magic_wand_Tablet_HIT_management_script_on_device	-0.12
Current_HIT_type_Speaker_dont_use	0.11
BetterSupportTabletManagement	-0.11
RateWatchManagement	0.10

Table 5.3: Each cluster along with their size and each two PCA component's mean value

PCA Component 1	**PCA Component 2**	**Cluster**	**Cluster Size**
-0.86	0.70	Cluster 1	66
-0.12	-1.32	Cluster 2	47
1.79	0.46	Cluster 3	35

Table 5.4: Each cluster along with their size and each PCA component's median value

PCA Component 1	**PCA Component 2**	**Cluster**	**Cluster Size**
-1.01	0.74	Cluster 1	66
-0.25	-1.23	Cluster 2	47
1.74	0.66	Cluster 3	35

features with meaningful contributions are identified, balancing significance, interpretability, and noise reduction. This threshold is low enough to capture important features and high enough to filter out minimal contributors, aligning with standard practice in exploratory analysis.

Following the Principal Component Analysis (PCA), we proceeded with k-means clustering to categorize our dataset. To identify the most suitable number of clusters, we employed the elbow method, the results of which are depicted in the elbow plot (see Figure 5.2). While the plot suggests an elbow point, a distinct potential elbow was noted at k=3. To validate the optimal number of clusters further, we also generated a silhouette score plot for a visual assessment of the clustering, illustrated in Figure 5.3. This visualization further supported the selection of k=3 as the optimal number of clusters, as it shows that all the cluster plots are more or less of similar thickness and hence are of similar sizes , underscoring superior cluster cohesion and separation with three clusters. Therefore, we selected 3 as the optimal number of clusters.

Subsequently, we describe each cluster. For that we used specialized approach as the following:

Overall Approach Explainability in machine learning and artificial intelligence remains a challenging problem. Each method of explaining model behavior has its strengths and weaknesses [221]. To address this issue comprehensively, we employed a multi-faceted approach, triangulating our explanation using three distinct methods: (1) Principal Component Analysis (PCA) (2) statistical differences and (3) decision tree. This combined approach allowed us to capture a more holistic and nuanced understanding of the crowdworker clusters.

Principal Component Analysis (PCA) We began by describing each cluster using both the mean and median values of the PCA components and the outcomes. PCA helped in reducing the dimensionality of our dataset, highlighting the most significant features that contribute to the variance within the data. By analyzing the central tendency measures of these components, we were able to characterize each cluster effectively. This step was crucial for understanding the underlying structure of the data and provided a foundation for further statistical analysis.

Statistical Differences Next, we examined the presence of significant differences among the clusters for each feature. Given that all the significant features were categorical, we applied the chi-square test to identify statistically significant differences. To ensure the robustness of our results, we adjusted the threshold values for significance using the Bonferroni correction method, which accounts for multiple comparisons. Features with statistically significant differences were further analyzed based on their effect sizes, employing a threshold of more than 0.3 to denote medium to large effect sizes. This method allowed us to quantify the magnitude of differences between clusters, facilitating a deeper understanding of their unique characteristics.

Decision Tree Finally, we employed decision tree analysis to understand the features that differentiate each cluster of crowdworkers. Decision trees provided a clear and interpretable model of how different features contributed to cluster membership. This step enabled us to identify the most critical features that distinguished each cluster, complementing the insights gained from the PCA and statistical tests. By integrating these three methods, we were able to triangulate our explanation, providing a comprehensive and nuanced understanding of the crowdworker clusters.

Before applying PCA, we aggregated similar tasks into broader categories: "completion tasks" (e.g., sentiment analysis, information finding, etc.) and "management tasks" (e.g., accepting tasks, communicating with requesters, etc.). This was done to reduce dimensionality, capture overall variance more effectively, and enhance interpretability. However, for decision tree analysis, we retained each specific task type as original features to leverage the granularity and assess feature importance accurately, allowing for precise classification and deeper insights into the factors driving the outcomes.

5.3 Result

In this section, we demonstrate the validity of our hypothesis concerning the systematic differences between crowdworker by detailing the clusters of crowdworkers based on the key characteristics of the work process including device type and stage of work. In addition to that we also discuss the distinctions among the cluster of crowdorkers based on these crucial aspects of their work process.

5.3.1 Description of each PCA Component

Here we present the findings from the principal component analysis (PCA), where each PCA component is described by the loading score of PCA component values.

PCA Component 1 – Preference of Non-workstation devices for both Completion and Management This PCA component as shown in Table 5.1 has a bucket of features focusing on better support for managing HITs on non-workstation devices including tablet phone, watch and speaker with absolute value spanning from 0.19-0.21.

The second group, with absolute values ranging from 0.10 to 0.18, shows less interest towards not using non-workstation devices for completing HITs in future and especilaly desire to use phone and tablet for completing survey type HITs. It highlights the versatility of using 'magic wand' for managing HIT functionalities across any device, such as script availability and filtering HITs on any device especially emphasizing tablet and phone. This component also reflects that the mentioned magic wand particularly would like to enhance the UX on any device type, pushing the boundaries of conventional workstation-based crowdwork.

Overall, this PCA component underscores the potential for improving HIT management and completion on non-workstation devices and highlights the innovative possibilities of supporting scripts and making UX bettter in enhancing usability across multiple devices.

PCA Component 2 – Selective Device Utilization for HIT Management and Completion This PCA component as shown in Table 5.2 has two buckets.

The first bucket, with feature values ranging from 0.18 to 0.20, indicates a positive rating for phones and tablets for both managing HITs. However there is a less enthusiam in using magic wand for supporting script on phones for managing HITs using phone.

The second bucket with absolute value spanning from 0.10-0.17 indicates a positive rating for tablet and phone for completing HITs. In addition to this, this bucket also reflect the reluctance to use watch, tablet and speaker for completing HITs. Additionally, this bucket reveals lesser enthusiasm for employing the 'magic wand' across all devices for supporting scripts to manage HITs epecially on tablet and also shows less enthusiam for better support in managing HITs using tablet.

Overall, this PCA component highlights a general preference for using phones and tablets for management and completion tasks, while indicating a lack of interest in supporting features like scripts or tools for managing HITs across any devices, particularly on phone and tablets.

5.3.2 Description of Clusters of Crowdworkers

Here, we describe the three clusters of workers based on their current work practices and preference of using devices on different stages of work as the following by correlating each PCA component based on their mean and median value for PCA based analysis, statistical difference and decision tree.

Cluster 1 – Traditionalists with Limited Use and Openness of Smartphone and Tablet

This cluster of 66 workers demonstrates a nuanced interplay of device preferences for completing and managing Human Intelligence Tasks (HITs), as captured by PCA components 1 and 2 as shown in Table 5.3 and Table 5.4 . PCA component 1, with its higher negative magnitude (mean = -0.86, median = -1.01), indicates a clear low enthusiasm for better supporting tablets, phones, watches, and speakers for management, also shown in Figure 5.4. For smartphones and tablets, this group of workers generally shows low enthusiasm for better supporting management tasks, particularly for accepting HITs on

Table 5.5: All the significant features along with their p-value within threshold effect size of more than 0.3 across all three clusters of workers

Feature	display	p-value	Effect-size	Corrected-effect-size
Better Support Tablet Management	$\chi^2(12, 148) = 89.1$	< 0.001	0.55	0.51
Better Support Phone Management	$\chi^2(12, 148) = 77.6$	< 0.001	0.51	0.47
Rate Tablet Management	$\chi^2(8, 148) = 70.9$	< 0.001	0.49	0.46
Better Support Speaker Management	$\chi^2(10, 148) = 69.0$	< 0.001	0.48	0.45
Better Support Watch Management	$\chi^2(12, 148) = 67.0$	< 0.001	0.48	0.43
Rate Tablet Completion	$\chi^2(8, 148) = 59.8$	< 0.001	0.45	0.42
Rate Phone Management	$\chi^2(8, 148) = 59.0$	< 0.001	0.45	0.42
Rate Phone Completion	$\chi^2(6, 148) = 48.9$	< 0.001	0.41	0.38
Desired HIT Type Tablet Survey	$\chi^2(2, 148) = 20.0$	0.007	0.37	0.35
Rate Watch Management	$\chi^2(6, 148) = 35.1$	< 0.001	0.34	0.32
Magic Wand Phone HIT Management Script on Device	$\chi^2(2, 148) = 17.1$	0.028	0.34	0.32

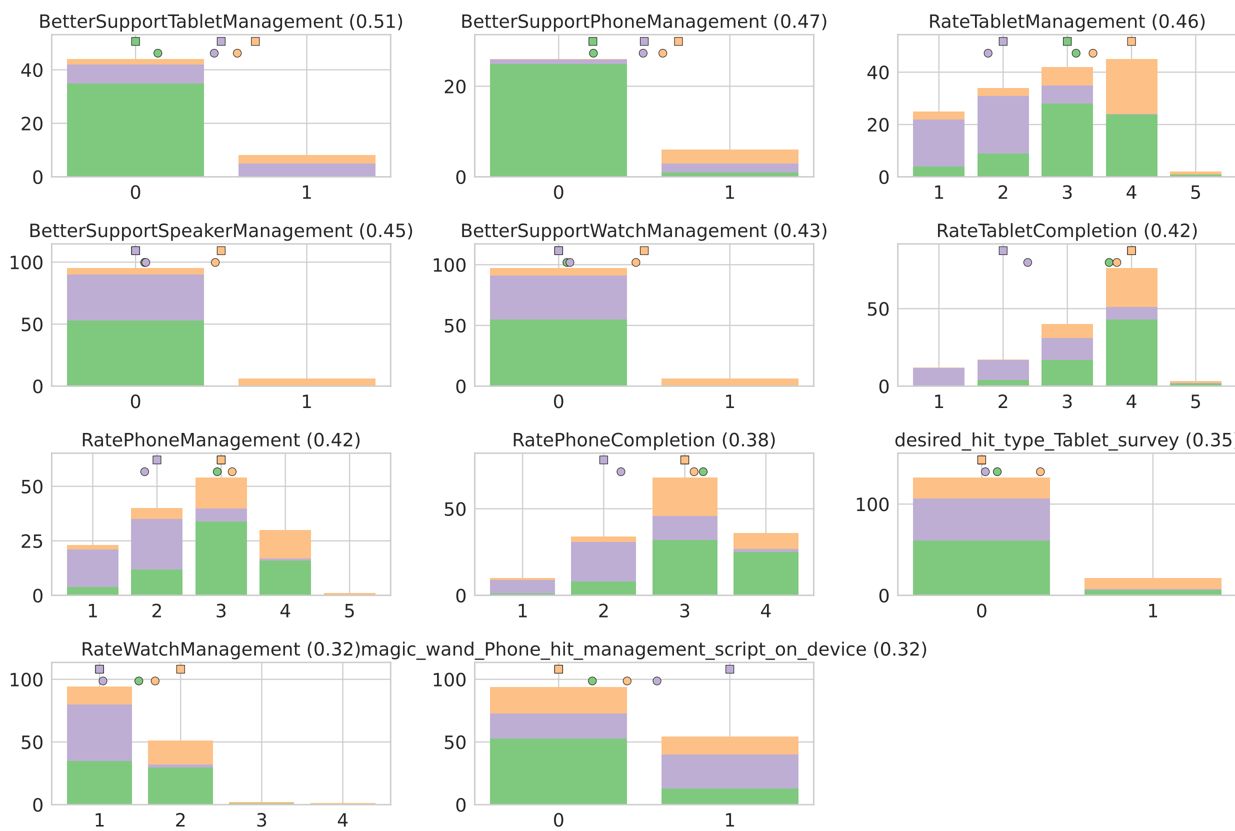


Figure 5.4: Visualization of all the significant features as shown in Table 5.5 within threshold effect size of more than 0.3 across all three clusters of workers

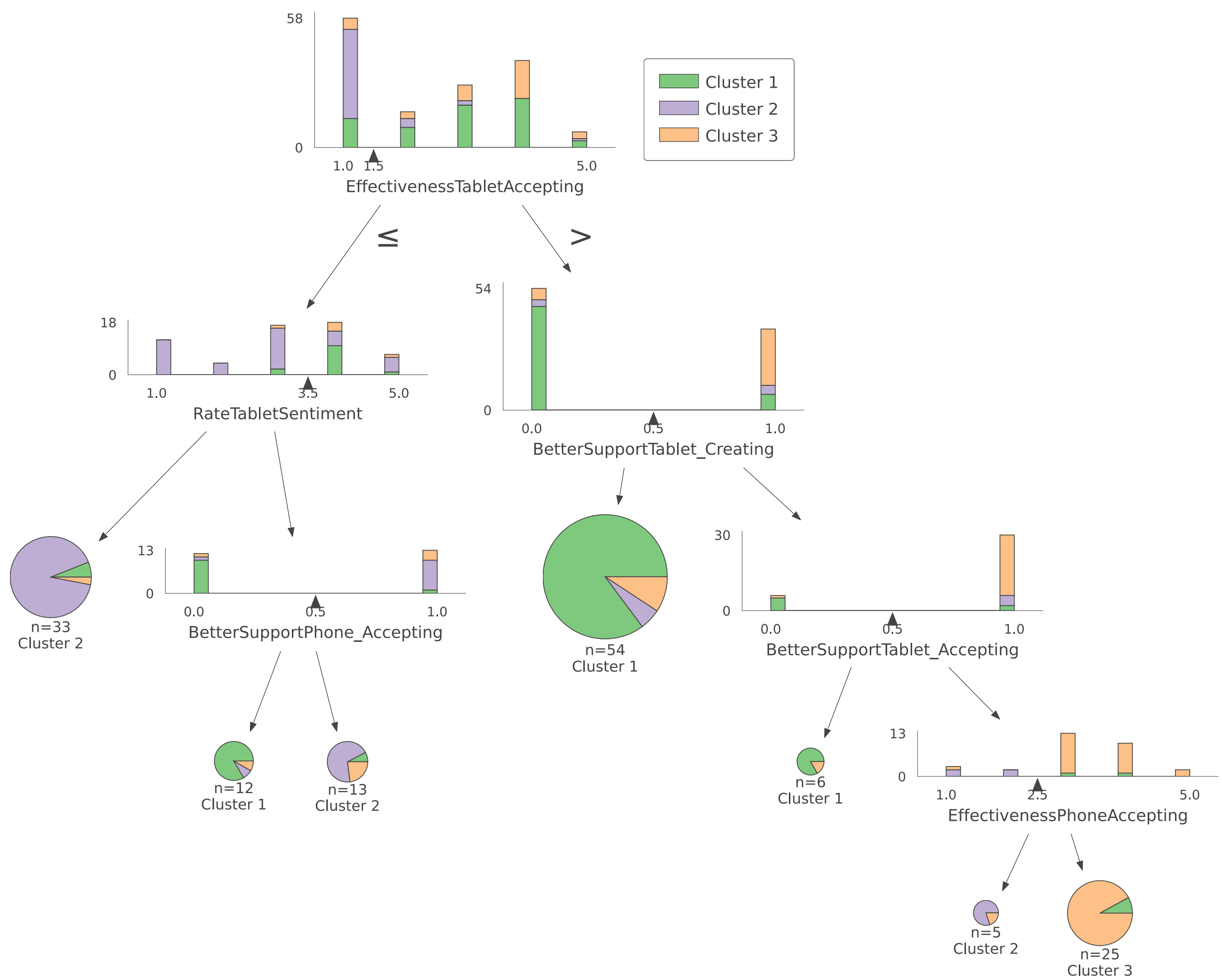


Figure 5.5: Decision Tree which shows all unrolled features that differentiate clusters of crowdworkers

phones and creating catchers on tablets, as illustrated in Figure 5.5. Additionally, these workers have no desire to complete HITs using any of these non-workstation devices. This sentiment is echoed in the following participant quotes:

> *"I honestly prefer to do all of my work on a laptop. I only use a smartphone if it is absolutely required for the HIT. So I don't really want to see any further support for it."*(P5)

> *" I don't care for working on my tablet."*(P73)

> *"I would not like to see any types of HITs better supported on the smart speaker"*(P70)

> *"I do not believe it would be possible to complete HITs on a smart watch"*(P99)

PCA component 2, which shows a positive direction (mean = 0.70, median = 0.74), reveals a favorable perception of phones and tablets for completing and managing related work. Figure 5.4 shows that the majority of this group rates tablets and phones highly, between 3 and 4 on a 5-point scale where 5 refers to most suitable, for managing and completing HITs. Despite these high ratings, they currently mostly use smartphone for those HITs which requires to use phone and tablets for completing survey HITs and only if the task requires to use tablet. Majority of the workers in this group do not use watches and speakers for completing HITs currently:

> *"Sometimes I will accept a HIT that needs to be completed on a smartphone . . . "*(P5)

> *"I do alot of surveys on my tablet."*(P35)

Despite some positive ratings for phones and tablets in specific scenarios, this group of workers prefer to avoid all non-workstation devices for completion and have less enthusiam in any better support for managing HITs on the non-workstation devices. This indicates a primary reliance on traditional work setting like workstation, with limited current use and openness of mobile devices including smartphone and tablet.

Cluster 2 –Traditionalist with Some Current use of Non-Workstation Devices and a Desire for Management Support mainly on Smartphone and Tablet

This cluster of 47 workers, characterized by their distinct device usage patterns for tasks, exhibits specific preferences as detailed in PCA components 1 and 2 as shown in Table 5.3 and Table 5.4 . PCA component 2, the more dominant with a higher magnitude and a negative direction, (mean = -1.32, median = -1.23) indicates a low rating for phone tablet for completing and managing HITs as well as watch for managing HITs . This is reflected in Figure 5.4 as the workers in this group rates tablets and phones between 1 and 2 on a 5-point scale for management and between 2-3 for completion of HITs where 1 refers to least suitable and 5 refers to most suitable.

They have some openness for using non-workstation devices currently for completion as they are currently using tablet, watch and speaker for completion and using phone for completing phone specific HITs :

> *"I only complete HITs which say they require a smartphone to complete . . . "*(P85)

> *"Generally easy ones on [tablet] if I'm away from my computer and I'm relaxing or watching TV."*(P11)

> *"I often use my smart speaker to get extra information to complete HITs"*(P33)

> *"I use [my watch] to bring in information to help with HIT's when they take up my whole screen."*(P32)

PCA2 also reflects that this group wants enhanced support features like scripts or tools on all devices, including workstations, phones, tablets, speakers, and watches, with a particular emphasis on phone and tablet as also shown in Figure 5.4. They also express a desire for better support in managing HITs on tablets:

> *"I would love to have an all in one script that allows me to accomplish all of my tasks unlike now where I need to have multiple scripts and screens open in order to accomplish my work [on workstation]"*(P7)

"Similar to a smartphone, I'd like to see more scripts and tools that help catch work on tablet"(P84)

"Make smart speaker easier to accept HITs and work on them or even find good work"(P124)

"I would make a HIT catcher for portables. To automatically catch HITs."(P99)

PCA component 1, though showing a lesser magnitude and negative direction (mean = -0.12, median = -0.25), corroborates reluctance of non-workstation devices for both completing and managing HITs , which implies the continued use of traditional workstations, highlighting a conventional approach to managing and completing HITs.

This cluster prefers traditional workstations for both managing and completing HITs highly. In addition to that they are currently using non-workstation devices like phones, tablets, watches, and speakers for specific tasks. The dominant sentiment is a desire for enhanced support features such as scripts and tools, particularly on phones and tablets, to facilitate HIT management better. This suggests a hybrid approach where traditional work setups are preferred, but there is a recognition of the potential benefits of using non-workstation devices for management with appropriate support tools.

Cluster 3 – Non-workstation Enthusiasts with Seamless HIT Management

This cluster of 35 workers, characterized by their distinct device usage patterns for tasks, exhibits specific preferences as detailed in PCA components 1 and 2 as shown in Table 5.3 and Table 5.4 .

PCA component 1, with the highest absolute magnitude and a positive direction (mean = 1.79, median = 1.74), indicates a strong preference for better support on all four studied non-workstation devices, especially phones and tablets, for managing HITs. This preference is also reflected in Figure 5.4. They also mention their high desire for incorporating non-workstation devices for completing HITs. This preference is highlighted by the following participant quotes:

"I really enjoy HITs where I have to speak phrases into my microphone, but those HITs are often not compatible for a smartphone, even though it would be much easier to do those types of HITs on a smartphone."(P12)

"All HITs should be better suited for a tablet . . . "(P33)

"Transcription HITs would be well-suited for a smart speaker . . . "(P21)

"I'd love to see more testing HITs on my smartwatch, like ones related to fitness or things like that. I love testing new apps, and I'd love to test ones for HITs on my smartwatch."(P12)

These workers also want support for scripts for managing HITs as well as filtering HITs on any device, especially on phones and tablets, and seek better HIT UX compatibility across all studied devices. This sentiment is further reflected in the following participant quotes:

"I would use my magic wand to create a more intuitive interface for completing HITs on a tablet . . . "(P22)

"I would add HIT catchers and scripts to run on the smartphone platform . . . "(P50)

"Having a version of the web version optimized for a tablet that allows for you to run scripts and tasks."(P55)

"I would setup tools for requesters to make HITs that can be completely answered using a smart speaker . . . "(P30)

"On the smartwatch, I would allow apps to find and catch HITs. Workers may be away from their workstations but can carry the watch with them, allowing them to catch HITs while they are away."(P61)

"I would change how easy it is to find HITs and organizing the tasks [on workstation]."(P55)

PCA component 2, though possessing less magnitude (mean = 0.46, median = 0.66) compared to PCA component 1, still contributes valuable insights. It shows that workers rate phones and tablets for both managing and completing tasks between 3-4 , where rate watches and speakers between 1-2 in a 5 point scale 5 represents high suitability and 1 represents low suitability, as shown in Figure 5.4. This cluster shows a strong preference for using non-workstation devices like phones and tablets for both completing and managing HITs. There is a clear need for greater flexibility and functionality on mobile and other non-workstation devices, emphasizing the importance of versatile and optimized interfaces.

In summary, the clusters differ in their acceptance and enthusiasm towards non-workstation devices for completion and management, with Cluster 1 showing the least enthusiasm, Cluster 2 being moderately open, currently use non-workstation devices along with with a desired support on smartphones and tablets for management, and Cluster 3 being the most receptive and enthusiastic about fully integrating non-workstation devices for managing and completing HITs.

5.4 Discussion

Our study shows that majority of the workers currently using workstation however they are open to using non-workstation devices and there is a small group who are enthusiastic about using only non-workstation devices for two stages of work. In this section we discuss the design guidelines based on our findings in crowdwork. We also discuss the unique work practices of these clusters of workers in contrast to traditional 9-5 white-collar jobs to understand crowdworkers' work practices with traditional workers.

5.4.1 Design Guidelines based on Clusters of Crowdworkers

Optimize Desktop Interfaces

Our findings show that the biggest cluster of crowdworkers indicates a strong preference for traditional work settings, focusing primarily on desktop usage. Therefore, it is important to prioritize robust functionality and intuitive navigation on desktop platforms, ensuring minimal latency and high performance for desktop-based tasks to maintain user satisfaction and productivity [186, 189].

For example, consider a crowdworker engaged in complex data analysis tasks. They rely on their desktop for its processing power and large screen to handle multiple data sets simultaneously. To support this, the platform should provide features like advanced data visualization tools, seamless multitasking capabilities, and a highly responsive interface. Minimal latency in loading data and executing commands ensures the worker can maintain a smooth workflow, thereby enhancing productivity and overall satisfaction. Prioritizing these aspects will help create an efficient and user-friendly desktop experience for crowdworkers.

Develop Hybrid Systems

The biggest cluster of crowdworkers (cluster 1) predominantly relies on traditional workstations, but they also show some openness towards using non-workstation devices, especially phones and tablets, for crowdwork. Therefore, it is essential to create systems that allow seamless transitions between desktops and non-workstation devices, ensuring tasks initiated on desktops can be easily continued on non-workstation devices like smartphones and tablets, and vice versa. This will help support the cognitive processes of workers distributed across different devices [205, 65].

For example, a crowdworker might begin their day by conducting detailed data entry on a desktop. Later, they may need to leave their workstation but want to continue working. With a seamless transition system in place, they can switch to their smartphone to review and edit the data entry while on the go. Features like cloud-based synchronization and a responsive design ensure that all progress is saved and the user interface remains consistent across devices. This capability allows the worker to maintain productivity and cognitive flow, regardless of the device they are using, enhancing overall efficiency and satisfaction.

Enhance Mobile-First Features and Implement Advanced Mobile Capabilities Our findings show that the non-workstation enthusiasts group of workers is currently utilizing and seeking enhanced support for non-workstation devices in both managing and completing HITs, aiming for greater efficiency and adaptability in their workflow, mostly on smartphones and tablets. Therefore, it is important to ensure that mobile interfaces are intuitive and easy to navigate by understanding the usability requirements from our chapter 4 along with previous work [210, 102] to enhance user experience and usability on the platforms.

For example, a crowdworker might use their smartphone to quickly browse and accept new tasks during a commute. An intuitive mobile interface with large, touch-friendly buttons and streamlined navigation helps them find and accept tasks with minimal effort. Additionally, implementing task management tools that enhance mobile device functionality, such as voice-activated commands, can further improve efficiency. Imagine a worker using voice commands to start a new task or check their task list while cooking dinner.

Moreover, ensuring mobile devices can handle complex tasks with multi-tasking support and enhanced processing power is crucial. For instance, a tablet might be used to simultaneously handle data entry and video reviews, with the platform dynamically adapting to the user's needs to provide a seamless experience. These innovations help workers maintain high productivity and adaptability in their workflow, leveraging the strengths of their preferred devices [118].

By focusing on these enhancements, platforms can significantly improve the user experience for non-workstation enthusiasts, ensuring they can efficiently manage and complete tasks on their preferred devices.

Incorporate Tools for a Smooth Workflow Across Different Devices

Our findings reflect an inclination for enhanced management on mobile devices, especially phones and tablets, from the second largest cluster of workers. Additionally, cluster 3 shows a preference for enhanced management across all devices, from workstations to non-workstation devices.

Therefore, it is important to provide multitasking management tools in crowdsourcing platforms that allow workers to manage and complete tasks efficiently on any device [141, 135]. For instance, a crowdworker might start their day on a desktop workstation, efficiently completing complex data analysis tasks. As they move through their day, they can switch to their smartphone to quickly accept new assignments or communicate with requesters using voice commands while waiting for an appointment. Later, they might use a tablet during a commute to review and organize tasks, utilizing a mobile-optimized interface that supports drag-and-drop functionality.

Introducing real-time collaboration tools optimized for specific device use, such as instant messaging and shared document editing on smartphones, and developing streamlined task-switching capabilities, can further facilitate smooth workflow transitions, supporting efficient and adaptable workflows [22, 49]. These features enable crowdworkers to maintain productivity and adapt to different work environments seamlessly, ensuring a cohesive and user-friendly experience across all devices.

5.4.2 Understanding Crowdworkers through the Lens of Traditional and Established Worker Groups

Our investigation has revealed patterns of interrelated variabilities concerning device type and work stage, including both task completion and management within the crowdwork community. This exploration into crowdworkers' practices, particularly when juxtaposed with the work habits of traditional 9-to-5 employees or other established workers, enriches our comprehension of contemporary work dynamics.

The predominant group of traditionalists mirror the conventional 9-to-5 work model, primarily relies on a single device type—echoing the traditional desk-bound employee—yet exhibits some openness towards integrating mobile devices in the workflow. This behavior resonates with the work habits of information workers, who, as noted in prior research [23], typically use workstations for completing tasks but also open towards using mobile devices [107, 136].

Workers within cluster 2 demonstrate a high degree of strategy, currently using non-workstation devices for some tasks for completion along with that they would like major support for scripts and tools for better management of HITs on all devices including workstation, phone, tablet, speaker and watch with a higher emphasis on phone and tablet. This strategic diversification of device usage not only optimizes their workflow but also aligns with the practices of modern information workers and IT professionals, who adeptly leverage the unique strengths of different technologies to enhance productivity and efficiency [72, 108, 165].

Cluster 3 workers, the non-workstation enthusiasts, exhibit work practices similar to digital nomads. Both groups prefer using mobile devices like smartphones and tablets for managing and completing tasks, valuing flexibility and mobility. This alignment is evident in their reliance on mobile technology for communication, project management, and accessing cloud-based tools [180, 153]. Both benefit from intuitive mobile interfaces, seamless device synchronization, and the ability to work from any location, emphasizing the need for versatile platforms that support these work styles.

Contrary to the inclinations of the 'Non-workstation Enthusiasts' cluster of crowdworkers, traditional information workers predominantly utilize desktops or laptops, supplemented by other devices to support their work [165, 162]. Non-workstation devices are primarily seen as auxiliary tools for tasks like automatic scheduling and receiving notifications rather than direct task completion and managment devices [5], the potential of these technologies to revolutionize communication, workflow [238], and overall worker well-being and productivity [4] is an area ripe for further exploration.

In conclusion, our study contributes to the discourse in HCI by unveiling a complex landscape of work practices that transcends the boundaries of traditional employment and adapts to the dynamic contours of contemporary work. By illuminating the diverse ways in which workers navigate and optimize their interactions with technology, our findings offer pivotal insights into the evolving nature of work, encouraging a re-imagined approach to the design and support of work practices in the digital age. This comprehensive understanding enriches the HCI community's efforts to craft technological solutions that are not only responsive to the current needs of the workforce but also anticipatory of the future trajectories of work.

5.5 Limitation

We acknowledge several limitations in this project. First, our participant pool is limited to 150 participants. While this sample size is adequate for the data collected and analyzed from MTurk, it restricts the generalizability of our findings. Future research should conduct large-scale quantitative studies to determine if these results are statistically generalizable to the broader population. Second, our research focuses exclusively on workers from Amazon Mechanical Turk. Although MTurk is a prominent platform for crowdwork, it is essential for future studies to examine whether these findings hold true across other crowdsourcing platforms, such as Prolific, CrowdFlower, and Upwork. Third, our results rely on self-reported data from crowdworkers. There is a possibility of participants misremembering details about their device usage, and response bias may lead them to present an overly positive view of their crowdwork practices. Consequently, the results should be interpreted with caution. This research should not be viewed as definitive; further studies are necessary to triangulate and validate our findings.

Chapter 6

Understanding platform-specific comparison of traits for supporting flexibility of crowdworkers

6.1 Introduction

In the rapidly evolving digital labor market, supporting the flexibility of crowdworkers is paramount. Diverse device types have become integral to effective workflows, and prior research underscores the importance of integrating different devices to enhance crowdworker flexibility [89]. However, each device type requires a tailored workflow to maximize efficiency.

Crowdworkers' standard workflows encompass two stages: managing tasks (such as finding, accepting, and communicating with requesters and peers) and completing tasks (including sentiment analysis, information finding, audio tagging, speech transcription, image classification, and bounding box annotation) [229, 159, 89, 33, 10, 58]. We hypothesize that the factors influencing task completion and management vary across different device types, requiring customized workflows. To investigate this, we already conducted a survey with 150 participants on Amazon Mechanical Turk (mTurk) to understand the factors and characteristics influencing task completion and management across different devices as shown in see chapter 3, chapter 4 and chapter 5.

Building on this initial research, we hypothesize that platform-specific factors could impact crowdworkers' device usage. Therefore, we extended our research from MTurk to Prolific by surveying 1000 workers and performing a comparative analysis. This comparative analysis aim to determine whether insights gained from MTurk (see Chapter 4, Chapter 5) are applicable to Prolific, thereby assessing if platform influence traits to support crowdworkers' flexibility. By conducting this survey on a larger scale, we aim to quantify not only the quantitative results but also the qualitative observations from mTurk to better support the flexibility of crowdworkers.

Additionally, we wanted to explore how different demographics impact current and desired usage of various device types. The demographic variables under consideration include general demographics (such as age, gender, cultural background), technical factors (e.g., device usage, interest in software and technology), and financial factors (such as income range, financial dependents) as well type of profession which have been shown to influence device ownership and usage patterns [171, 134].

In this chapter, we aim to address four research questions toward the goal of understanding the impact of demopgraphics on crowdworkers device usage in work:

1. RQ1. Is there a correlation between demographic factors and device practices for completing and managing crowdwork?

2. RQ2. Is there a correlation between technical factors and device practices for completing and managing crowdwork?

3. RQ3. Is there a correlation between profession outside of crowdwork and device practices for completing and managing crowdwork?

4. RQ4. Is there a correlation between financial factors and device practices for completing and managing crowdwork?

Our key findings include,

1. Our findings indicate that workstations are the most commonly used devices for crowdwork, followed by smartphone for both current and desired use. Although the

current usage of smart speaker and smartwatche is nonexistent on both MTurk and Prolific, there is notable interest among Prolific workers in using these devices for future task completion. Both MTurk and Prolific workers express some interest in incorporating these devices for task management.

2. Both MTurk and Prolific workers desire using smartphones for finding and accepting tasks. Additionally, MTurk workers prefer smartphone and tablet for creating and listening to catchers, while Prolific workers primarily prefer smartphone for receiving notifications. On MTurk, smart speaker and smartwatch are favored for listening to catchers and finding tasks, respectively. Conversely, Prolific crowdworkers prefer these device types for receiving notifications.

3. Our findings indicate that both MTurk and Prolific crowdworkers predominantly use desktops or workstations for task completion and management. However, among non-workstation devices, MTurk workers favor smartphones and tablets the most. Prolific, on the other hand, has a more diverse cluster of workers with six different device usage patterns compared to MTurk's three clusters. This indicates a broader range of device preferences among Prolific crowdworkers, including greater interest in integrating tablets, smartwatches, and smart speakers.

4. Our findings show a correlation between crowdworkers' gender and region, and their use of desktops and smartphones in crowdwork. Additionally, there is a correlation between workers' comfort with using known technology and their interest in using smart speakers for managing crowdwork in an ideal scenario. Moreover, desktop usage outside of crowdwork correlates with its use for crowdwork tasks both currently and in ideal conditions with high-quality tooling. Furthermore, having a profession outside of crowdwork is correlated with a preference for using smart speakers to manage crowdwork.

These key findings provide a comprehensive understanding of current practices, preferences, and potential areas for improvement in crowdsourcing platforms, ensuring better alignment with crowdworkers' needs and enhancing overall platform effectiveness to support the flexiblity of crowdworkers.

6.2 Study Design

In our research, we conducted a survey of primarily 1,000 Prolific workers to explore crowdworkers' perceptions and desires regarding crowdwork. After manually reviewing and approving additional responses that were not initially accepted by Prolific, we included a total of 1,004 participants. In this section, we describe the survey design, the execution of the study, and the methods used to analyze the collected data. Our IRB-approved survey includes 26 questions across six sections. The full survey is available in the Appendix B.

Section 1: General Demographics In this section, we collected information on participants' educational backgrounds, including degrees and areas of specialization, and their self-described ethnicity and cultural backgrounds, allowing for rich, personalized insights. Additionally, the Prolific platform automatically captured data on age, gender, country of residence, and primary language.

Section 2: Technological Demographics In this section, we asked participants about the types of devices they own, offering choices such as desktop or laptop, smartphone, tablet, smart speaker, and smartwatch. Additionally, we inquired about the frequency of their daily device usage, extending beyond crowdwork-related activities. To gauge participants' technical proficiency, we asked them to rate their agreement with statements regarding their enthusiasm for experimenting with new devices and software, as well as their preference for using familiar technologies.

Section 3: Crowdwork Demographics This section asks about crowdworkers' demographics such as their tool usage on crowdwork, the types of crowdsourcing platforms they use during crowdwork besides Prolific, the amount of hours they work per week to complete and manage crowdwork, as well as the number of years they have been working as a crowdworker.

Section 4: Financial Demographics In this section, we asked crowdworkers about financial aspects, including the percentage of their total income derived from crowdwork, their financial situation compared to others in their region (with options ranging from far below average to far above average), and the number of dependents they are responsible for financially supporting.

Section 5: Device Usage In Section 5, we asked questions about completing and managing crowdwork using the devices they own. We explored their frequency of using owned devices for these tasks and the importance of contextual factors like place and time. We also investigated challenges preventing device use, such as lack of tools, poor formatting for the device, explicit prohibition of device use, difficulty in managing crowdwork on the device, and other challenges, as well as no challenges. Participants were also asked to imagine an ideal world with high-quality tooling available for completing and managing crowdwork on each device they own and to indicate their frequency of using each device type in such a scenario. Additionally, we assessed their interest in completing different types of crowdwork (e.g., surveys, testing software, transcription, classification, information finding) and managing crowdwork tasks (e.g., finding and accepting work, communicating with requesters, receiving notifications). All the given options were informed by our previous work in Chapter 4. Although the ideal scenario question was asked on MTurk in Chapter 4 and does not reflect actual device usage, it remains valuable. It provides a baseline for understanding device usage preferences in an ideal context, offering insights into potential improvements and future directions for enhancing crowdwork platforms.

Section 6: Crowdworkers' Opinions In this section, we gathered crowdworkers' opinions on several factors impacting their work experience. We asked about their ability to do crowdwork at their chosen time and place, on their preferred device, with ample time to complete tasks, and the availability of various types of crowdwork. We also explored the importance of various platform features when deciding which crowdsourcing platform to use. These features included support for non-desktop devices, effective communication between workers and requesters, sufficient platform features, prevention of bots, fair compensation, fair treatment of workers and requesters, ease of finding work, a strong community of crowdworkers, protections for privacy and personal data, and the ability to make an impact in the world.

These questions were inspired by our previous work in Chapter 4 and previous research [116, 43, 232].

6.2.1 Survey Development

After creating an initial version of our survey, we improved it over several rounds of internal review from our team. Once we produced a version we were satisfied with, we submitted for and received IRB approval for our study. We piloted the survey with a group of 30 individuals. Finding no substantial issues, we proceeded to launch the finalized survey.

6.2.2 Study Execution

After being approved by our institution's IRB, we deployed our survey as a task on Prolific on April 16, 2024. Participants were compensated USD $1.75 for completing the study. The average completion time was approximately 5-10 minutes resulting in an average compensation of $12/hour and was deployed to 8 regions including USA, Canad, UK, Europe, Asia, Africa, Americas, Oceania.

6.2.3 Analysis

The multiple-choice survey questions were analyzed using quantitative methods, specifically through the computation and reporting of response frequencies. For the few open-ended response questions espcially cultural background as well as questions where other option were provided, we analyzed the data using a methodology inspired by grounded theory [204]. Throughout the qualitative process, we used the constant comparative methods [64] to identify codes that were originally separate and could be combined, as well as codes that were originally combined and should be split.

Second, after open coding was finished, the coders conducted axial coding, grouping related concepts, generating themes, and describing how the concepts related to the themes and each other. We do not continue through selective coding (the final step of grounded theory). Through this process, the two coders kept a detailed set of research notes. These notes aided in the process of coding but also included insights and lessons learned as the coders completed the coding process. According to grounded theory, these notes are often just as valuable as the actual codes. Many of our findings are contextualized based on the insights found in the research notes.

In the results section, the reported percentages for the general demographics (e.g., education, cultural background), crowdworker demographics, financial demographics, and crowdworker opinions are based on responses from all 1004 participants, as each participant provided responses for these categories. For the technical demographics, the type of device used outside of crowdwork and technical interest are also based on responses from all 1004 participants. However, the device usage frequency within technical demographics is based on the number of participants who own each specific device type. Similarly, in the device usage section, all reported percentages for each device type are based on the number of participants who own each respective device type.

6.3 Understanding Prolific Crowdworkers from their Response of Survey Questions

In this section, we discuss each part of the survey, including general demographics, technological demographics, crowdworkers' work experiences, financial demographics, and device usage in crowdwork. We also examine crowdworkers' opinions on flexibility in time, place, and device usage, as well as their views on the features supported by crowdsourcing platforms. This analysis aims to understand the nature of Prolific crowdworkers, which will help the community more effectively leverage this platform in the future. Additionally, it provides context for the remainder of our results to understand the type of crowdworkers based on their device usage and stage of work along with understanding impact of demographics on crowdworkers device usage in work practces.

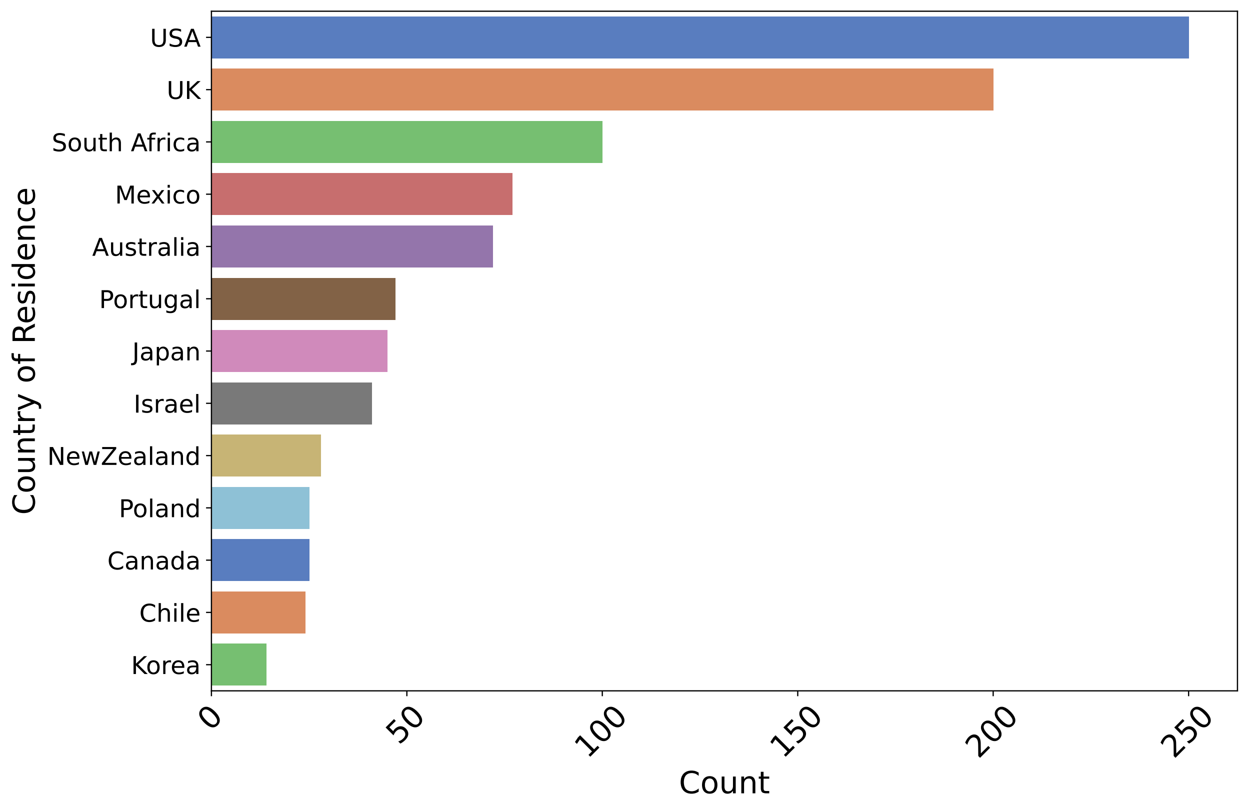


Figure 6.1: Visualization of distribution of residence of participants

6.3.1 General Demographics

Most of the participants belong to the age group of 25-44 (n = 663, 66.03%). Nearly half of the participants were between the ages of 25 and 34 (43%). More than half of the participants are female as (n = 534, 53.2%).

As the majority participants were assigned from USA and UK, natually most of the participants (n = 550, 54.78%)reside in the USA or UK. Other significant regions include Mexico and Australia, as shown in Figure 6.1. Since most participants reside in English-speaking countries, it is natural that the primary language for most participants is English (n = 650, 65.74%). Additionally, about 10% of participants (n = 105) have Spanish as their primary language.

Among all participants, the majority self-describe their ethnicity or cultural background as white, white British, or white European, with 3.39% identifying them as black. Most participants held a bachelor's degree (38.7%, n = 389) or an advanced degree (23.3%, n = 234), with only 1.1% (n = 11) not completing high school. Finally, participants, on average, had 734 approved tasks.

6.3.2 Technical Demographics

Regarding the devices they owned, nearly all participants had a workstation (n = 948, 94.42%) and a smartphone (n = 944, 94.02%). The majority also owned a tablet (n = 524, 52.19%), while fewer participants owned a smart speaker (n = 303, 30.18%) or a smartwatch (n = 371, 36.95%). Among participants who owned these devices, the majority used them outside of crowdwork very often to always: desktop (n = 838, 83.47%) and smartphone (n = 899, 89.54%). In contrast, very few people used smartwatches (n = 242, 24.10%), tablets (n = 214, 21.31%), and smart speakers (n = 121, 12.05%) at a high frequency (see Figure 6.3).

Regarding technology interest, the majority of participants (n = 904, 90.03%) showed a interest in trying known technology. Additionally, a similar number of participants (n = 897, 89.34%) also expressed interest in trying new technology. This indicates that most participants not only enjoy familiar technology but are also curious to try out new innovations (see Figure 6.4).

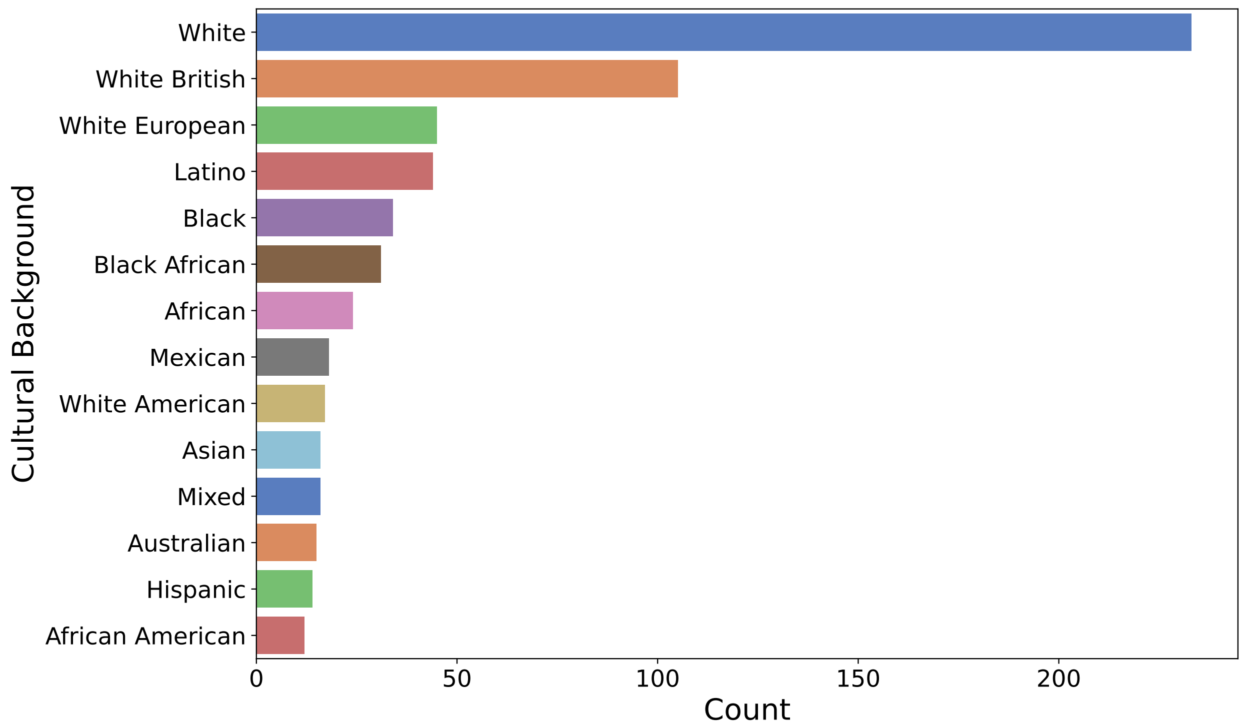


Figure 6.2: Visualization of distribution of mostly mentioned cultural background of participants

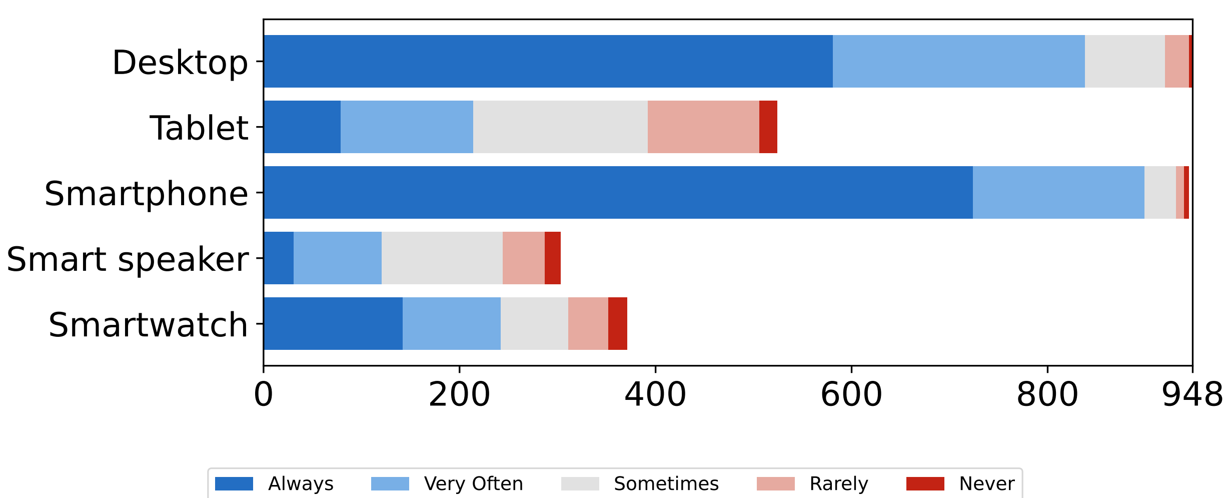


Figure 6.3: Visualization of device usage frequency for day to day tasks outside of crowdwork

6.3.3 Crowdworkers Demographics

Other than Prolific, Amazon Mechanical Turk (n = 111, 11.05%) and Upwork (n = 104, 10.35%) are the next most used platforms by participants (see Table 6.1).

The majority of participants (n = 861, 85.76%) mentioned that they do not use any tools or scripts to do crowdwork. This indicates a potential issue with tool awareness or no requirements to use tools to manage or to complete crowdwork on this platform. Among the tools mentioned, MTurk Suite (n = 64, 6.37%) and Turker View (n = 27, 2.69%) are the most frequently mentioned tools. Although "Prolific Assistant" is a known tool for Prolific, only 2.19% paraticipants mentioned about using it. Most of the participants who use tools for crowdwork use only a single tool (see Figure 4.10 and Figure 4.11).

The majority of participants are new to crowdworking (see Figure 6.6). On average, participants invested 9.5 hours (+/- 9.5 hours; 95% CI) per week completing tasks (see Figure 6.7) and 1.61 hours per week managing tasks (see Figure 6.8). This indicates that participants, on average, invest more time completing crowdwork tasks rather than managing them.

6.3.4 Financial Demographics

For most of our participants, crowdwork acts as supplemental income, not their primary source of income. On average, participants reported that crowdwork makes up roughly a fifth of their income (18%) as shown in Figure 6.9. This implies that most crowdworkers have a separate income source other than crowdwork. Most participants hold IT or sevice position outside of crowdwork (see Figure 6.10. Most of this profession outside crowdwork is fully in person (n = 438 , 43.62%). In terms of financial dependents, most participants (n = 539, 53.7%) do not have financial dependents.

6.3.5 Crowdworkers' Device Usage

Here we present the findings related to participants' current device usage practices, as well as their device usage practices in an ideal world where high-quality tooling is available for each device type.

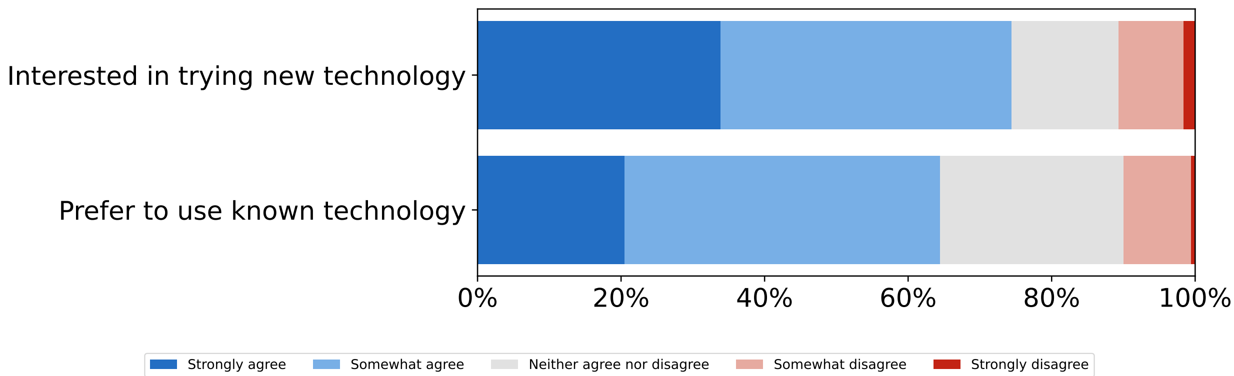


Figure 6.4: Visualization of agreement in interest in technology by participants

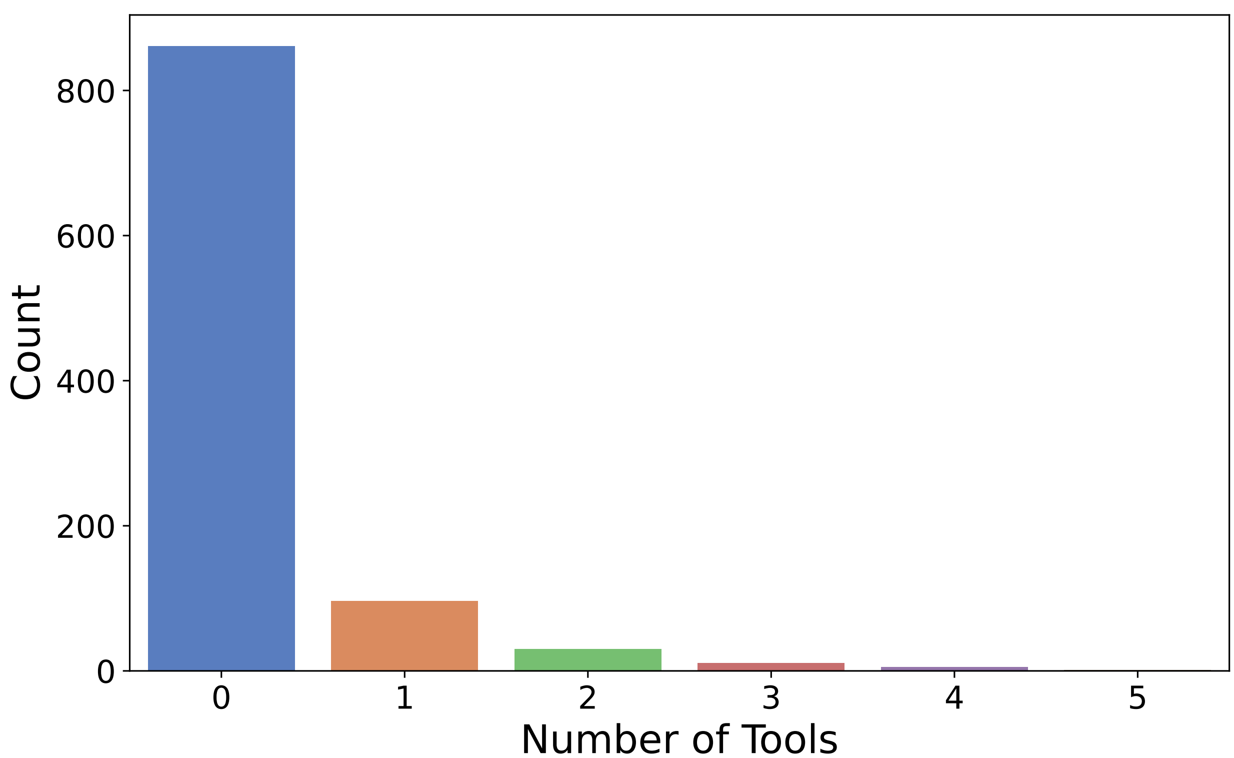


Figure 6.5: Visualization of no of tools used by participants

Table 6.1: Usage distribution of various crowdsourcing platforms mentioned by participants

Platform	Count
Amazon Mechanical Turk	111 (11.05%)
Upwork	104 (10.35%)
Other platforms	89 (8.86%)
CrowdFlower	9 (0.90%)
InnoCentive	5 (0.50%)
Topcoder	4 (0.40%)
HeroX	1 (0.09%)

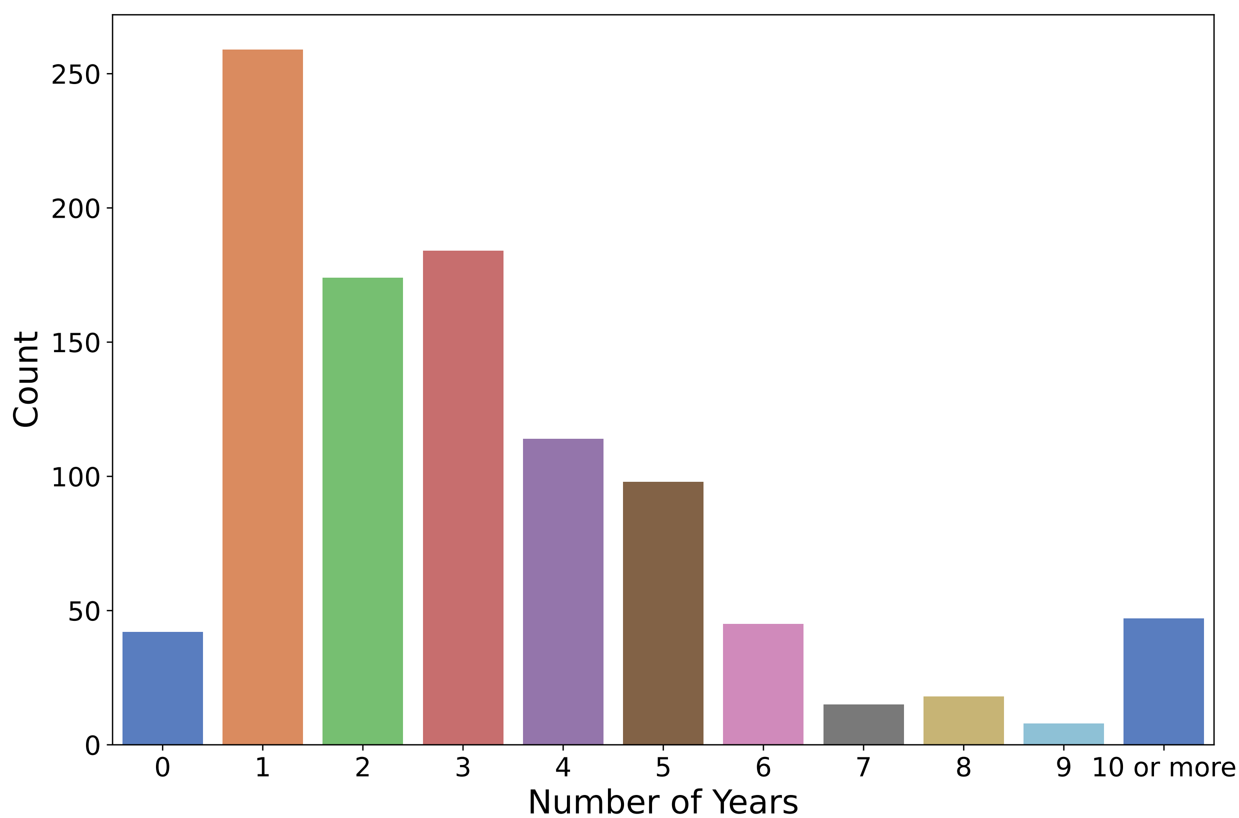


Figure 6.6: Visualization of no of years participants working as a crowdworker

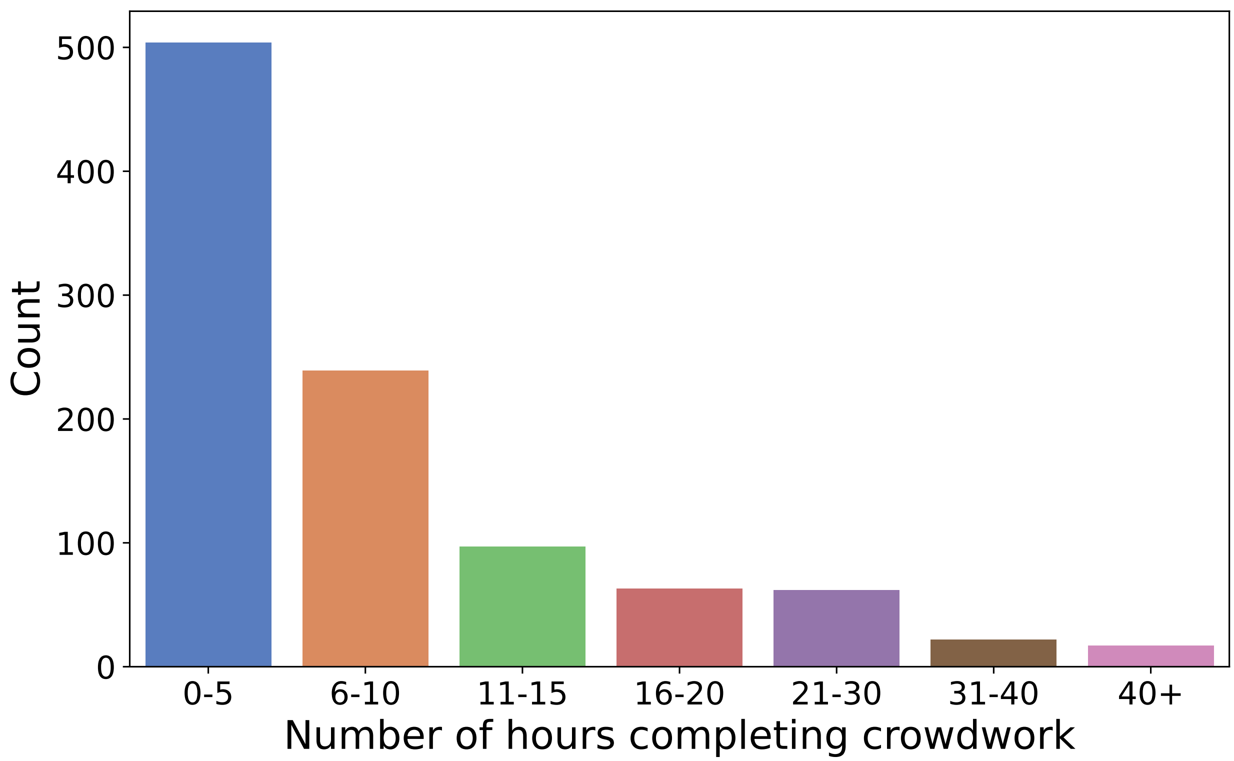


Figure 6.7: Visualization of no of hours participants use to complete crowdwork

Table 6.2: Usage Distribution of Tools

Tool	Count
Not using any tool	861 (85.76%)
MTurk Suite	64 (6.37%)
Turker View	27 (2.69%)
HIT Forker	26 (2.59%)
Other tool	22 (2.19%)
Panda Crazy	20 (1.99%)
Turk Guru	17 (1.69%)
Stax	18 (1.79%)
Otto	11 (1.10%)
Quebicle	9 (0.90%)

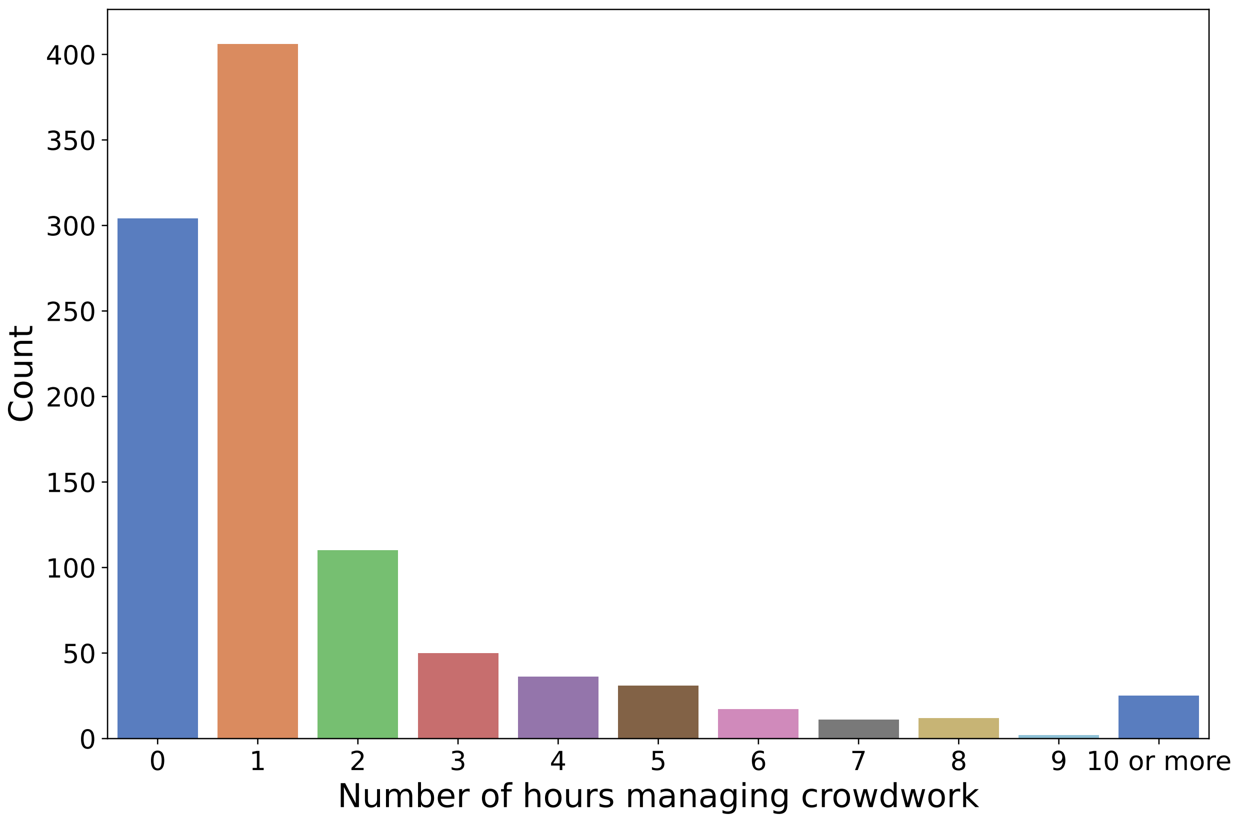


Figure 6.8: Visualization of no of hours participants use to manage crowdwork

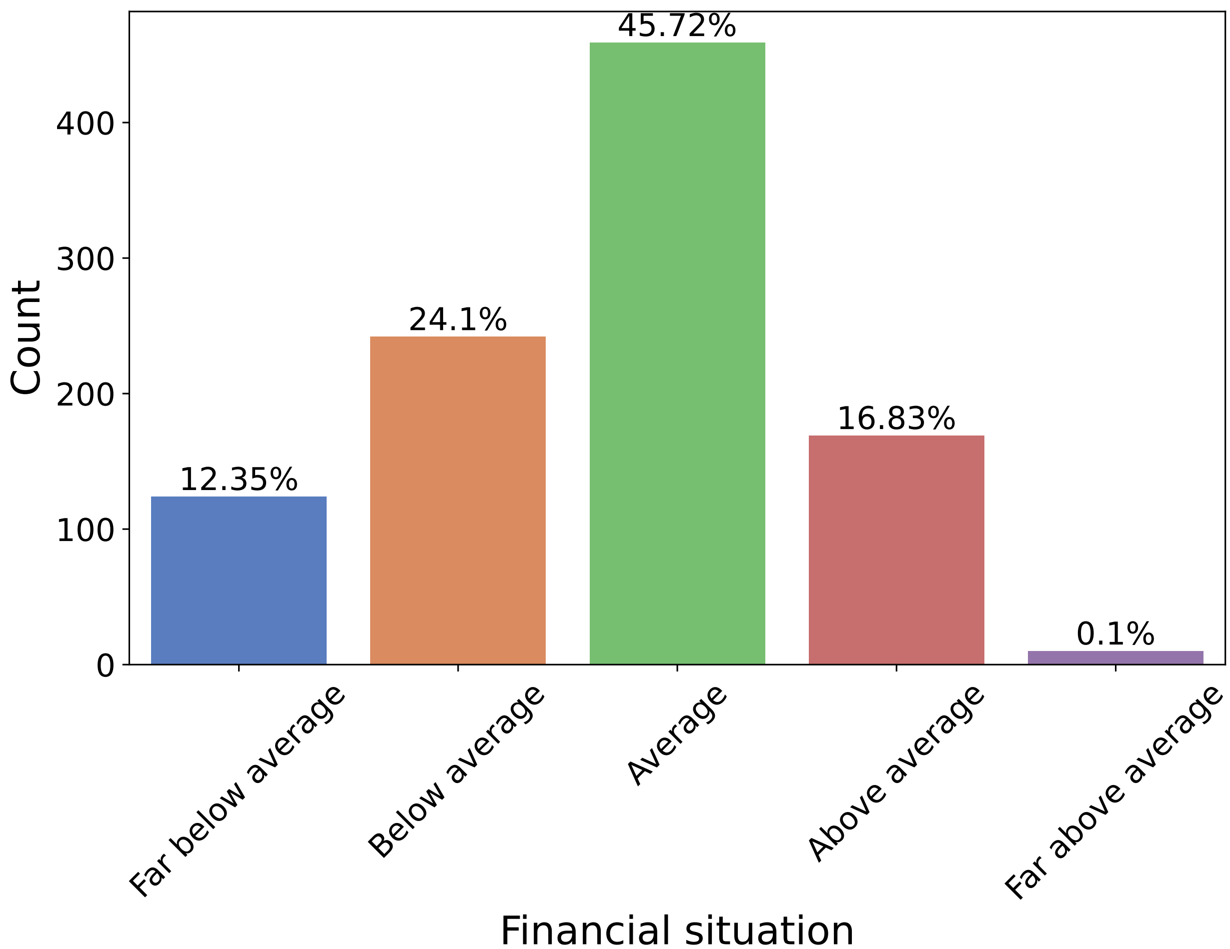


Figure 6.9: Visualization of distribution of financial situation mentioned by participants

Current Device Usage

In terms of current device usage, we find that nearly every participant (n = 882, 93.03%) primarily uses a desktop to complete crowdwork. Around a quarter of participants at least at least occasionally use a tablet (n = 187, 35.68%) and a smartphone (n = 587, 58.46%) to complete crowdwork. While completing crowdwork is rare for smart speakers and smartwatches, it is not entirely non-existent, as shown in Figure 6.11.

Device usage for managing crowdwork is similar to device usage for completing crowdwork. Most participants (n = 656, 69.20%) primarily use their desktop for crowdwork management, although fewer people use a desktop for management than for completion (see Figure 6.12). 26% participants (n = 139) use a tablet to manage crowdwork at least sometimes, whereas 47% of participants (n = 447) use a smartphone for management. Management of crowdwork is rare, but not non-existent, for smart speakers and smartwatches.

Contextual factors like place and time play an important role in the use of desktops, smartphones, and tablets for completing and managing crowdwork (see Figure 6.13 and Figure 6.14). The importance of place and time is almost non-existent for smart speakers and smartwatches. This indicates that since most participants currently do not use speakers and watches for crowdwork, they do not consider the importance of contextual factor.

Participants also mentioned challenges that prevent them from using devices as much as they would like for crowdwork as shown in Table 6.3. For desktops, the most frequently mentioned challenge is that a lot of crowdwork explicitly prohibits their use as shown in and not being able to use it when outside of workplace or commuting (n = 21, 2.21%) For tablets and smartphones, most participants cited issues with the UX/UI, stating that crowdwork is not well-formatted for these devices. Additionally, they mentioned that many tasks explicitly prohibit the use of smartphones and tablets. In the case of smart speakers and smartwatches, the most mentioned challenges include crowdwork not being formatted for speakers and the difficulty of effectively managing tasks using a speaker.

Ideal world Device Usage

In the current crowdwork ecosystem, most participants feel they need to use a desktop to manage and complete crowdwork due to the lack of support for other device types or other challenges. We are not sure whether the situation would be the same if there were full support for crowdwork on non-desktop devices. Therefore, we inquired whether participants would have the same or different opinions when high-quality tooling is available for each device type.

We see high interest in using desktop after high quality. In addition to that, there is significant demand for better support for the completion of crowdwork on smartphones and tablets. We find that after high-quality tooling support, most participants would like to use a smartphone (n = 868, 91.95%) to complete crowdwork, whereas 68% of the participants (n = 358) would at least occasionally like to use a tablet. A sixth of participants would like to use smart speakers (n = 46, 15.18%) and a seventh would like to use smartwatches (n = 55, 14.82%) who own them (see Figure 6.15).

For the completion of various types of crowdwork, we see high interest in using desktops irrespective of the types of crowdwork. In the case of non-workstation devices, there is high interest in completing surveys along with 35% workers interested in information finding tasks for tablets. A similar pattern is seen for smartphones, along with completing classification tasks. Most workers are not interested in using speakers and smartwatches for completing various types of crowdwork, people who are most of them prefer speaker for completing transcription related tasks and smartwatches for survey as shown in Figure 6.16.

Similarly, in the case of managing crowdwork in an ideal world, in addition to desktop there is also significant demand for better support for managing crowdwork on smartphones and tablets (see Figure 4.13). Sixty percent of participants would like to use a tablet (n = 315) and 70% would like to use a smartphone (n = 665), whereas one-eighth of participants (n = 35, 11.55%) would like to use speakers and one-ninth of participants who own the device type (n = 59, 15.90%) would like to use smartwatches. This shows that smartphones, tablets, and speakers are preferred more for completion than management, whereas smartwatches are favored more for management than completion.

Similarly like completion desktop is preferred the most for various types of managing crowdwork. In the case of non-workstation devices, tablets and smartphones show high interest, with 39.0%(n = 160) and 65.0% (n = 515)of respondents very interested, respectively for receiving notifications. Along with that tablet (39%, n = 160) and smartphone (64%, n = 510) also preferred for finding and accepting tasks as shown in Figure 6.18.

Overall, the findings highlight desktops as consistently the most preferred device. Smartphones generally serve as a secondary preference, offering flexibility and connectivity. Tablets show a mixed response across the tasks. While not as preferred as desktops or smartphones, tablets still play a crucial role, particularly in somewhat interested categories, indicating that this device type serves as a supplementary device rather than a primary one in crowdwork settings. In contrast, smart speakers and smartwatches,are unsuitable for crowdwork.

6.3.6 Crowdworkers Opinions

Here we discuss the findings regarding crowdworkers' opinions on completing and managing crowdwork at their preferred time, place, and device, as well as the importance of various features related not only to the management and completion of crowdwork but also to ethics, privacy, and community support provided by the crowdsourcing platform.

Opinions in Flexibility of choosing Time, Place, Device and Variery of Task

Most participants place high importance on choosing time (n = 875, 87.15%) and place (n = 820, 81.67%) along with the ability to have ample time to complete crowdwork(n = 780, 77.69%). Additionally, they value the ability to choose their own device type (n = 763, 76%) as well as variety of work (n = 736, 73.31%) as shown in Figure 6.19. This reflects that while most participants currently use desktops, they prefer the flexibility to choose their own time and place for crowdwork. This flexibility ensures they do not miss work opportunities when they are away from home or on the move as this is mentioned in the challenges for using desktop. Furthermore, being able to choose their own device type and task type would

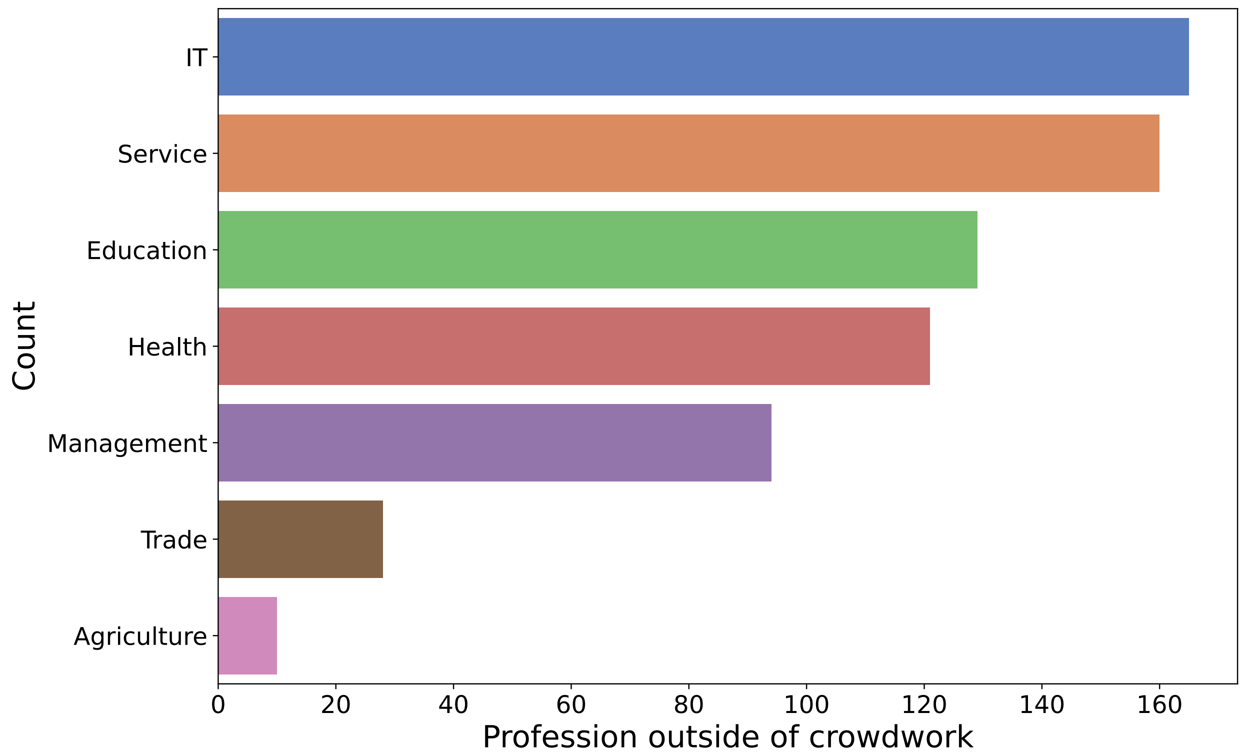


Figure 6.10: Visualization of distribution of profession outside of crowdwork

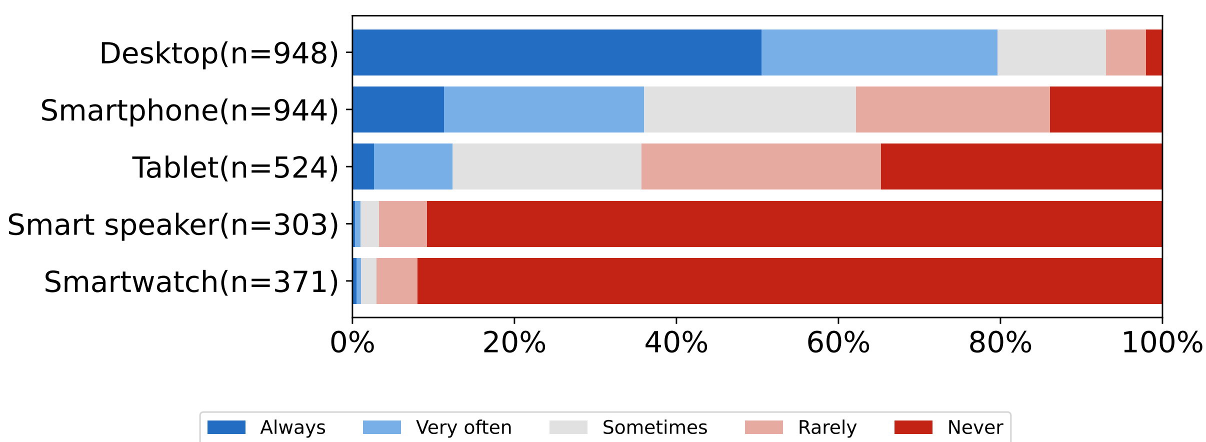


Figure 6.11: Visualization of distribution of current device usage for completing crowdwork

Table 6.3: Distribution of challenges mentioned by participants when using different devices for crowdwork

Challenge	Desktop	Tablet	Smartphone	Smart Speaker	Smartwatch
Crowdwork not well formatted	85 (9.0%)	230 (43.9%)	521 (55.2%)	153 (50.5%)	226 (60.9%)
Lack of tools	85 (9.0%)	137 (26.1%)	289 (30.6%)	141 (46.5%)	203 (54.7%)
Explicitly prohibits use of device	105 (11.1%)	186 (35.5%)	497 (52.6%)	121 (39.9%)	187 (50.4%)
Difficult to effectively manage crowdwork	58 (6.1%)	170 (32.4%)	326 (34.5%)	146 (48.2%)	230 (62.0%)
Other challenges	92 (9.7%)	21 (4.0%)	15 (1.6%)	13 (4.3%)	16 (4.3%)

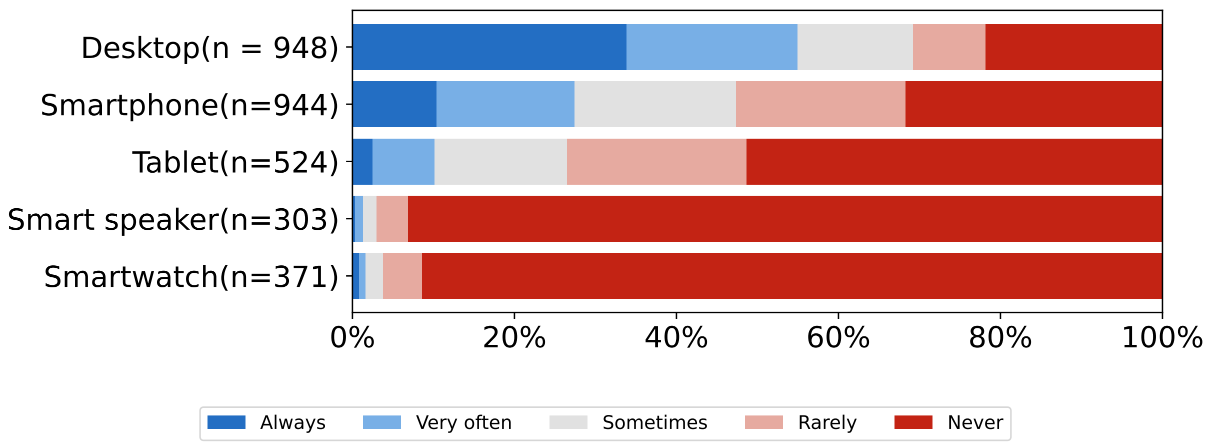


Figure 6.12: Visualization of distribution of current device usage for managing crowdwork

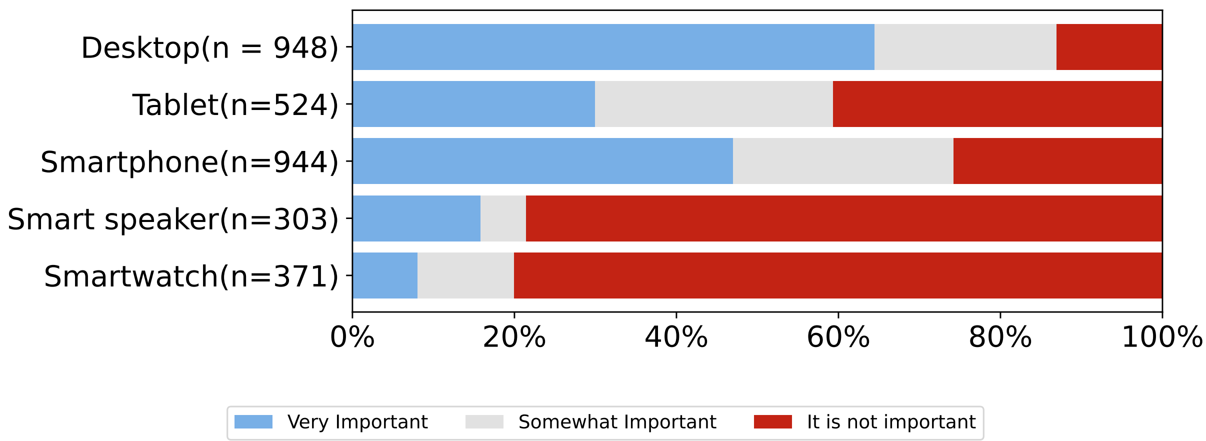


Figure 6.13: Visualization of importance of place for device usage in crowdwork

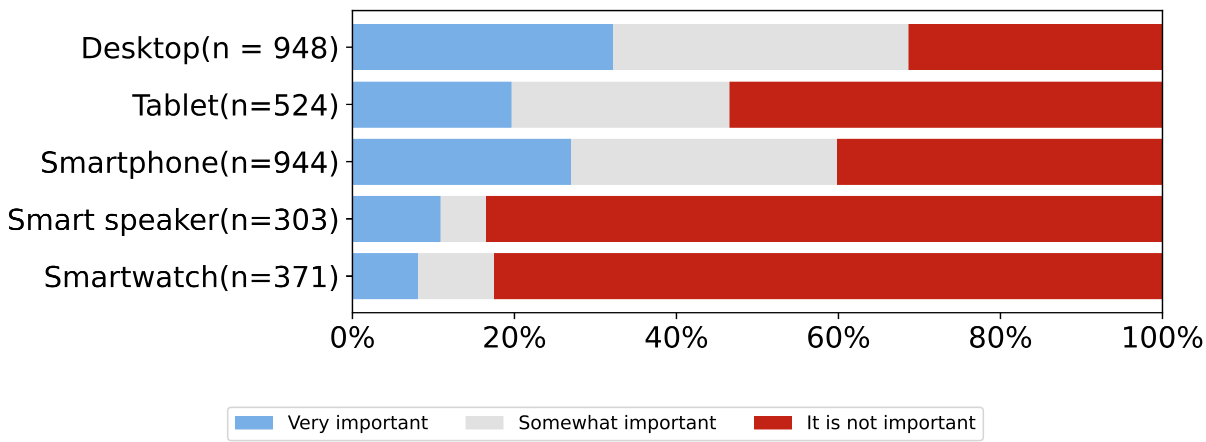


Figure 6.14: Visualization of importance of time for device usage in crowdwork

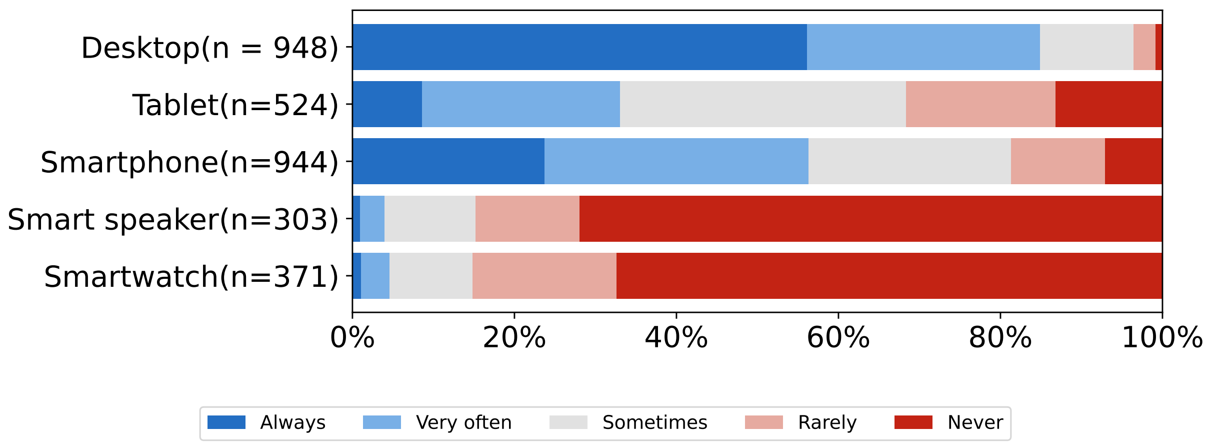


Figure 6.15: Visualization of frequency of device usage in an ideal world for completing crowdwork

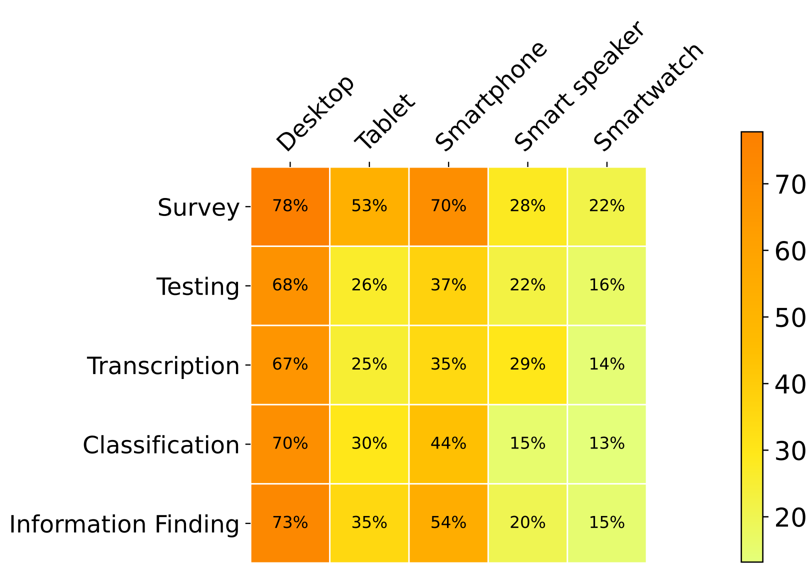


Figure 6.16: High interest of completing five different types of tasks including survey, testing, transcription, classification and information finding across studied devices by participants

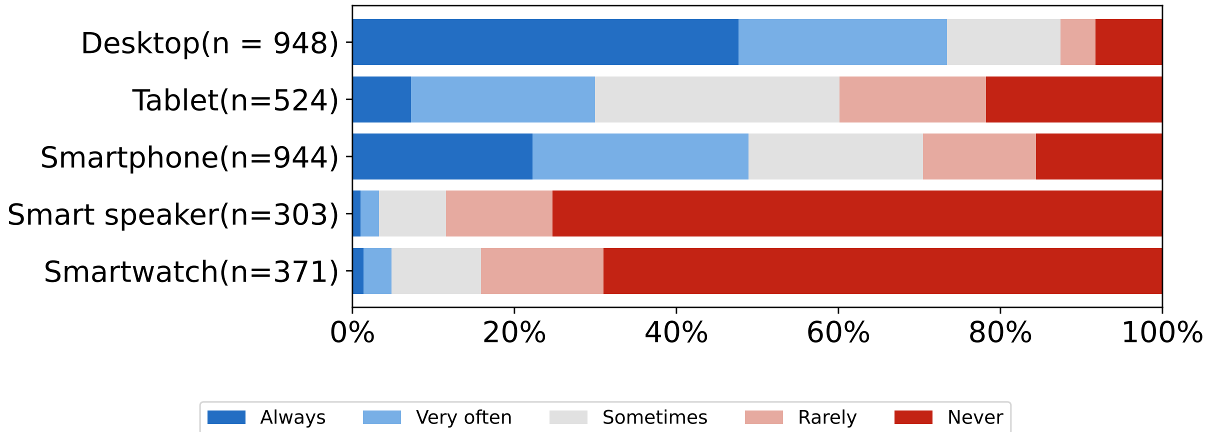


Figure 6.17: Visualization of frequency of device usage in an ideal world for managing crowdwork

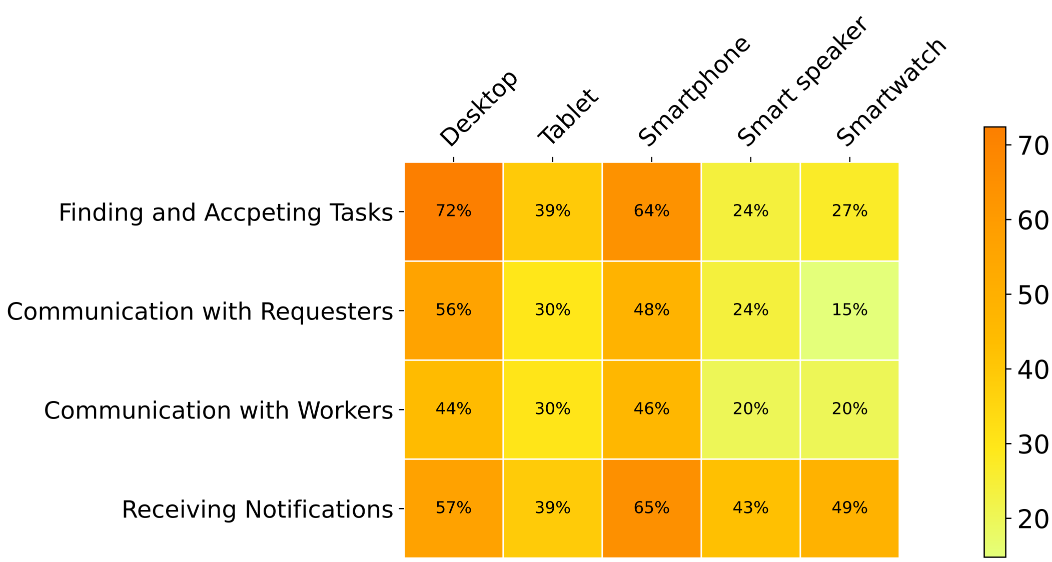


Figure 6.18: High interest of managing four different types of tasks including finding and accepting, communication with requesters and workers as well as receiving notifications, across studied devices by participants

allow them to complete and manage the type of work they want at any preferred time and place, enhancing their ability to manage and complete work efficiently.

Opinions of Participants in Crowdsourcing Platform Supported Features

Here we discuss crowdworkers opinion in platform supported features that focus on supporting features related to completion and managment as well as features the support the overall crowdwork ecosystem as shown in Figure 6.20 and Figure 6.21.

Crowdwork Management and Completion Related Features: Our finding reveals several key factors (refer to Figure 6.20) that crowdworkers prioritize in platforms. A significant majority of participants (n = 931, 92.73%) emphasize the importance of the platform making it easy to find crowdwork opportunities. A substantial portion of crowdworkers (n = 710, 70.72%) also appreciate platforms that provide sufficient features, eliminating the need for additional tools and scripts. This indicates that the majority of workers would like to first focus on features that ensure they get enough work on a day-to-day basis to raise their income level.

Participants also value support for communication between requesters and workers (n = 637, 63.45%), suggesting that clear, effective communication channels are an important management factor for crowdworkers.

In addition, around 54% (n = 544) of participants indicated the high importance of platform support for completing tasks on non-desktop devices, while 52.19% (n = 524) emphasized the need for support in task management on these devices. This indicates that more than half of the participants have a strong preference for platforms that offer robust compatibility for non-desktop devices.

Crowdwork Ecosystem Related Features: Our findings (see Figure 6.21) reveals that participants prioritize platforms that ensure fair compensation (n = 907, 90.34%) and fair treatment for workers (n = 900, 89.64%). This underscores a strong desire among crowdworkers for equitable and just working conditions. They also highly value measures to prevent bots from stealing work from legitimate crowdworkers (n = 893, 88.95%).

Additionally, there is a high value placed on platforms that offer robust protections for privacy and personal data (n = 880, 87.65%) and support the fair treatment of requesters (n = 844, 84.06%). This highlights the critical need for a secure and fair working environment that safeguards the interests of both crowdworkers and requesters.

Moreover, participants also emphasize the importance of providing impactful work (n = 534, 53.19%) and fostering a strong community of crowdworkers (n = 516, 51.39%). These findings suggest that beyond fairness and security, there is a notable interest in meaningful work and a supportive community among crowdworkers.

Overall, these findings suggest that to enhance the flexibility of crowdworkers, platforms need to prioritize flexibility in device usage, integrate comprehensive task management features, and uphold high standards of fairness and security. Addressing these areas will likely lead to a more fleixble, effective and appealing platform for crowdworkers.

6.4 Comparison of Overall Findings Between Prolific and Amazon MTurk

In this section, we discuss the similarities and differences between overall Prolific results and Amazon MTurk results. Understanding these platform-specific nuances is crucial for interpreting research findings accurately and ensuring their applicability across different crowdsourcing environments. The exploration of platform biases, particularly comparing MTurk and Prolific, reveals distinct differences in how each platform operates and influences crowdwork practices. On MTurk, task design constraints are generally more flexible, allowing for a broader range of task types and designs. This flexibility facilitates the rapid acquisition of large samples but can lead to varying levels of task quality and participant engagement. MTurk's diverse and extensive workforce requires researchers to implement additional screening measures to ensure high data quality.

In contrast, Prolific enforces stricter participant recruitment criteria and offers more detailed screening options. This results in a higher-quality, more specific participant pool and greater transparency regarding participant demographics. Researchers on Prolific can more easily target specific populations, ensuring a diverse and representative sample. However, the stricter criteria may constrain certain types of tasks or research projects.

Now, let's delve into the similarities and differences between the overall results from Prolific and Amazon MTurk.

6.4.1 General Demographics

The gender distribution between MTurk and Prolific exhibits notable differences. On MTurk, 65% of participants are male, while 35% are female. Conversely, Prolific has a higher percentage of female participants at 53.2%, with 43.6% male and 3.2% opting not to disclose their gender. Age distribution is somewhat similar, with the majority of participants on both platforms falling within the 25-44 age range: 71.62% on MTurk and 64.44% on Prolific. However, Prolific attracts a more highly educated population, featuring a higher proportion of participants with graduate degrees (23.3%) and a lower percentage of participants whose highest education is high school(10.6%) compared with MTurk (see Table 4.1).

6.4.2 Owned Device Type

Device ownership trends show that workstations are the most commonly owned devices on both MTurk (99%)and Prolific(around 94%), followed by smartphones(MTurk = 98%, Prolific = 94%) and tablets(MTurk = 84%, Prolific = around 52%). Although the sequence is similar , more people own each of the above mentioned device type in MTurk than Prolific. A notable difference is observed in the ownership of other smart devices. On MTurk, more participants own smart speakers (82%)compared to smartwatches(74%), whereas on Prolific, smartwatches (around 37%) are slightly more common than smart speakers(around 30%). smart speaker than watch

6.4.3 Tool Usage

Tool usage shows significant differences between the platforms. On MTurk, every participant uses some type of tool to aid in their tasks, with MTurk Suite being the most commonly used tool as shown in Figure 4.10. In contrast, a significant majority of Prolific users (85.76%) do not use any tools. Popular tools on MTurk include Panda Crazy and HIT Forker besides MTurk suite, while on Prolific, people who use tool MTurk suite is the most common tool (6%)Turker View and HIT Forker have much lower usage rates around 3% . MTurk users are more likely to use multiple tools, with 47% using two or more (see Figure 4.11), compared to the low tool adoption on Prolific as shown in Figure 6.5.

6.4.4 Current device Usage Frequency for Management and Completion

Both platforms currently using desktops for completing and managing tasks. Neither platform shows significant use of smart speakers or smartwatches for these purposes and both these platforms currently use more smartphone than tablet for completion and as shown in Figure 4.1 and Figure 4.7 for MTurk and Figure 6.11 than Figure 6.12.

6.4.5 Desired Task Completion on Non-Desktop Devices

When it comes to task completion on non-desktop devices, preferences have similarities as well as they vary between the platforms as shown in Figure 4.12 and Figure 4.12:

- MTurk workers primarily prefer tablets for testing and surveys and smartphones for survey and software testing, whereas Prolific users also prefer both tablet and smartphones mainly for completing surveys, other than that also they also prefer these devices for completing information-finding tasks .
- Majority of crowdworkers on both platforms do not prefer using smart speakers and smartwatches for completing tasks. However among those who are interested, they prefer this device type for completing survey tasks for smart speakers on MTurk, while Prolific workers prefer this device type mainly for completing transcription and survey tasks.
- Both MTurk and Prolific users desire to use smartwatches for completing survey.

6.4.6 Desired Task Management on Non-Desktop Devices

Non-desktop device usage for task management also shows some similarities as well as distinct patterns as shown in Figure 6.18 , Figure 4.13 and Figure 4.14 :

- On MTurk, tablets are preferred for finding and accepting tasks and creating and listening to catchers, while on Prolific, they are similarly mainly used for finding and accepting tasks along with receiving notifications.
- MTurk workers prefer smartphones for active task management such as finding , accepting as well as creating and listening to catchers, while Prolific workers prefer this device type the most for receiving notifications. Besides that they also prefer to use this device type for finding and accepting tasks.
- MTurk workers prefer smart speakers for listening to catchers and accepting tasks , along with that few percentage mentioned their preference for receiving notifications. On contrary, Prolific workers prefer this device type primarily for receiving notifications.
- On MTurk, smartwatches are preferred for task acceptance and finding along with receiving notifications, while on Prolific, this device type is mainly preferred for receiving notifications.

Additionally. Prolific workers also value high importance on able to find work easily , sufficient features on platform itself that prevents external tool usage, commuication between requesters and workers. Over half of the participants place high importance on platform supporting completion and mangament of work on non-desktop devices as shown in Figure 6.20

6.4.7 Challenges with Device Use

Challenges reported by participants highlight platform-specific issues. On MTurk, 51% of participants mentioned about advanced tools to manage work. UI/UX issues are significant, particularly for smartphones and tablets as shown in Figure 4.14. On Prolific, lack of tools for managing tasks is a major concern across all devices. UI/UX challenges are significant for non-desktop devices, with a high percentage of participants reporting issues as shown in Table 6.3

6.4.8 Features related to Crowdwork Ecosystem

Both platforms encounter challenges within their ecosystems as shown in results section of Chapter 4 and Figure 6.19 and Figure 6.21, particularly in preventing bot activity and ensuring fair wages.

Additionally, Prolific users also place high value on equitable treatment for both workers and requesters and the importance of robust privacy and data protection measures.

In summary, MTurk and Prolific attract different demographics and exhibit distinct patterns in device ownership, tool usage, task completion, and management. MTurk participants are more likely to use multiple tools and engage in active task management, while Prolific users prefer passive engagement and focus on notifications. Both platforms face similar ecosystem issues, whereas Prolific users also place a high importance on fair treatment, privacy, and data protection. These differences reflect the unique user bases and operational dynamics of each platform, influencing how tasks are completed and managed by crowdworkers.

6.5 Clusters of crowdworkers in Prolific

6.5.1 Analysis

In this case we used the same analysis stragety as mentioned in Chapter 5. Before applying PCA, we aggregated similar tasks into broader categories: "completion tasks" (e.g., survey, iclassification, etc.) and "management tasks" (e.g., finding and accepting tasks, communicating with requesters and workers, etc.). This was done to reduce dimensionality, capture overall variance more effectively, and enhance interpretability. However, for decision tree analysis, we retained each specific task type as original features to leverage the granularity and assess feature importance accurately, allowing for precise classification and deeper insights into the factors driving the outcomes.

In our case we selected PCA components as five as shown in Figure 6.22. In case of choosing clusters, we chose the cluster no as 6 based on the elbow plot (see Figure 6.23) as well as based on the silhouette score plot illustrated in Figure 6.24. After identifying features

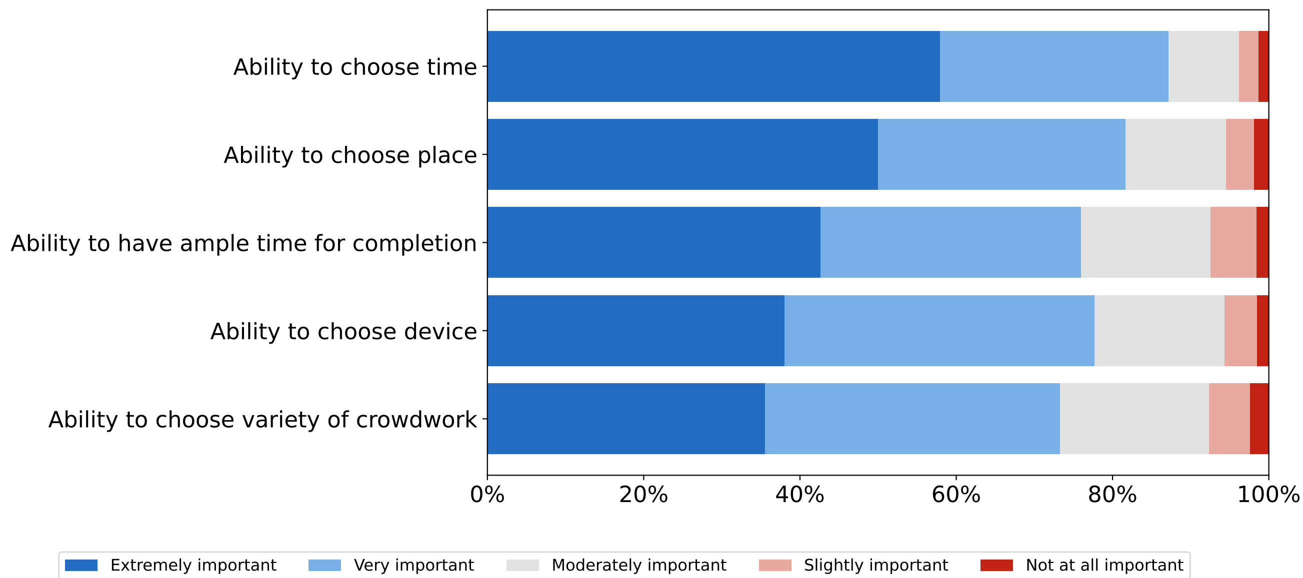


Figure 6.19: Visualization of importance of various factors when managing and completing crowdwork including time, place, device and variety of task according to participants

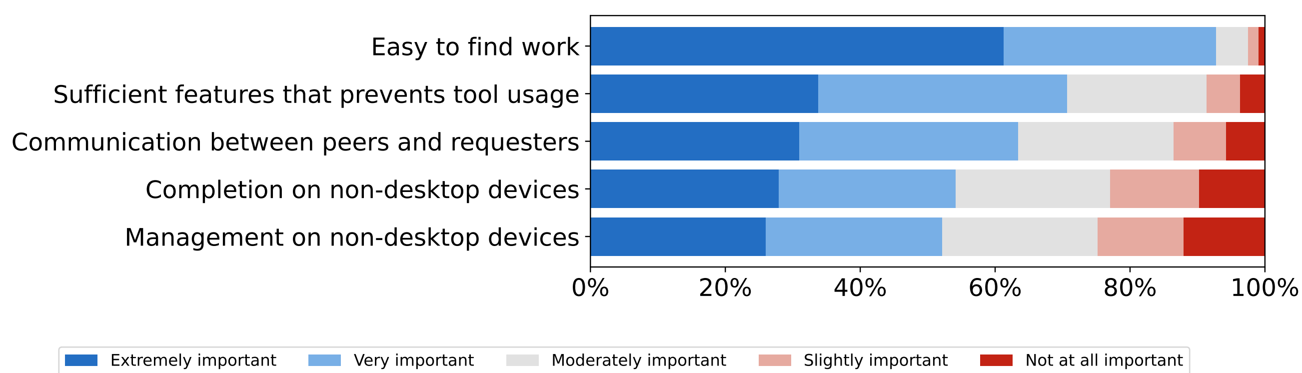


Figure 6.20: Visualization of importance of platform supported features related to completion and managment of crowdwork to participants

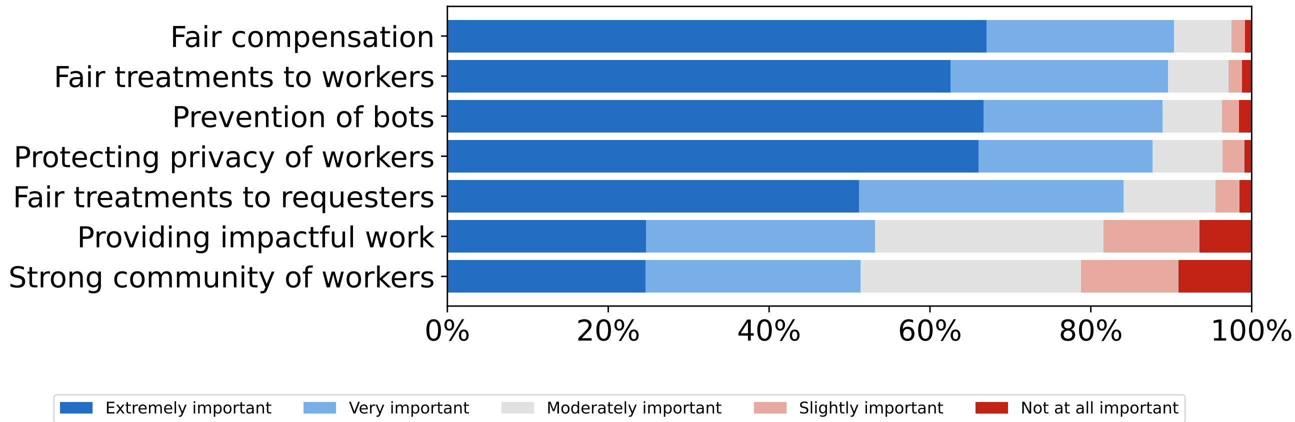


Figure 6.21: Visualization of importance of platform supported features realted to crowdwork ecosystem to participants

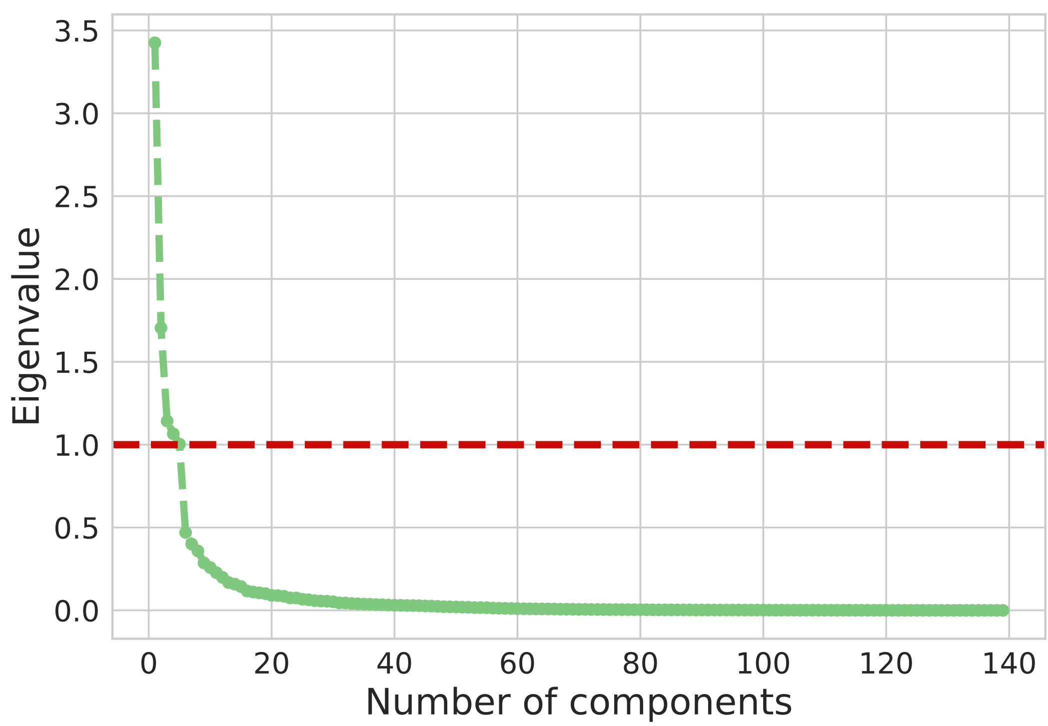


Figure 6.22: Scree Plot which shows the optimized number of PCA components 5 for our dataset

with statistically significant differences across clusters, we analyzed the distinctions between clusters based on their effect sizes, employing a threshold of more than 0.5 to denote a large effect size to keep a high filter as lots of features showed up with effect size more than 0.3.

6.5.2 Result

Here we demonstrate the cluster of crowdworkers as well as the difference between them to validate and generalize the result found from Amazon Mturk in chapter 5. We discuss the compariosn of clusters of crowdworkers between Mturk and Prolific in Section 6.4.

6.5.3 Description of each PCA Component

Here we present the findings from the principal component analysis (PCA), where each PCA component is described by the loading score of PCA component values.

PCA Component 1 –Dynamic Device Utilization with High Focus on Tablet

This PCA component analysis delves into the nuanced device usage patterns among crowdworkers, focusing on how various devices are utilized for completing and managing crowdwork and the associated challenges and preferences. This component consists of three buckets with their loading score positive value as shown in Table 6.4.

The first bucket consisting of features has the absolute value of PCA loading score from 0.27-0.42. This segment highlights the widespread ownership and favorable perception of tablets among workers. The data suggests that tablets are preferred for both completing and managing various types of crowdwork in an ideal setup, underscoring their versatility. Contextual factors like location play a significant role in using tablets, more so than temporal factors. Furthermore, there is a positive assessment of current tablet use for managing and completing crowdwork, indicating effective integration of this device into the workers' routines.

The second bucket contains all the features with the absolute value of PCA loading score from 0.10-0.26. This section reveals that smartwatches and smart speakers are commonly owned, but their use in crowdwork is fraught with challenges. The data shows that

Table 6.4: Top features of PCA Component 1 along with their loading factor within threshold absolute value 0.10

Features	**Loading Factor**
Owned device Type Tablet	0.42
TabletGivenCompletionCrowdwork	0.40
TabletGivenManagmentCrowdwork	0.39
TabletUsagePlaceImportance	0.37
No Tablet owned	-0.37
TabletUsageTimeImportance	0.35
IdealWorldTabletCompleteCrowdwork	0.34
UsageTablet	0.33
IdealWorldTabletManageCrowdwork	0.33
TabletCompleteCrowdwork	0.28
TabletManageCrowdwork	0.27
Owned Device Type Smartwatch	0.22
No Smartwatch owned	-0.20
UsageSmartwatch	0.19
Prohibits use of tablet	0.19
Owned device type Speaker	0.18
SmartwatchUsagePlaceImportance	0.16
Difficult to manage work on Watch	0.16
SmartwatchUsageTimeImportance	0.16
No Speaker owned	-0.16
Difficult to manage work on tablet	0.15
Lack of tools on tablet	0.15
Lack of tools on watch	0.15
UsageSmartspeaker	0.14
Prohibits use of watch	0.14
SmartphoneGivenManagmentCrowdwork	0.13
SmartspeakerUsagePlaceImportance	0.13
SmartspeakerUsageTimeImportance	0.13
IdealWorldSmartwatchManageCrowdwork	0.13
IdealWorldSmartwatchCompleteCrowdwork	0.13
Lack of tools on speaker	0.12
SmartwatchManageCrowdwork	0.12
SmartwatchCompleteCrowdwork	0.12
Difficult to manage work on speaker	0.11
Prohibits use of speaker	0.10
IdealWorldSmartspeakerCompleteCrowdwork	0.10
SmartspeakerManageCrowdwork	0.10
IdealWorldSmartspeakerManageCrowdwork	0.10
SmartphoneGivenCompletionCrowdwork	0.10
IdealWorldSmartphoneManageCrowdwork	0.10

Table 6.5: Top features of PCA Component 2 along with their loading factor within threshold absolute value 0.10

Features	**Loading Factor**
Owned device type Smartwatch	0.29
No Smartwatch owned	-0.28
UsageSmartwatch	0.25
Difficult to manage watch	0.25
SmartwatchUsagePlaceImportance	0.22
Lack of tools on Smartwatch	0.22
SmartwatchUsageTimeImportance	0.22
Prohibits Use of Smartwatch	0.20
Owned Device Type Tablet	-0.19
IdealWorldSmartwatchManageCrowdwork	0.18
IdealWorldSmartwatchCompleteCrowdwork	0.18
No Smartspeaker owned	-0.17
TabletUsagePlaceImportance	-0.16
TabletGivenCompletionCrowdwork	-0.16
SmartwatchManageCrowdwork	0.15
UsageTabletoutsideofCrowdwork	-0.15
SmartwatchCompleteCrowdwork	0.15
TabletUsageTimeImportance	-0.15
No tablet owned	0.15
WatchGivenManagmentCrowdwork	0.15
IdealWorldTabletCompleteCrowdwork	-0.15
TabletGivenManagmentCrowdwork	-0.14
WatchGivenCompletionCrowdwork	0.14
IdealWorldTabletManageCrowdwork	-0.14
Difficult to manage work on speaker	0.14
UsageSmartspeakeroutsideofCrowdwork	0.13
SmartphoneGivenManagmentCrowdwork	0.13
Lack of tools on speaker	0.13
TabletCompleteCrowdwork	-0.13
SmartspeakerUsagePlaceImportance	0.13
SmartspeakerUsageTimeImportance	0.12
TabletManageCrowdwork	-0.12
Prohibits use of speaker	0.11
SmartphoneGivenCompletionCrowdwork	0.11
IdealWorldSmartspeakerCompleteCrowdwork	0.10
IdealWorldSmartspeakerManageCrowdwork	0.10
SmartspeakerManageCrowdwork	0.10

Table 6.6: Top features of PCA Component 3 along with their loading factor within threshold absolute value 0.10

Features	**Loading Factor**
SmartphoneGivenManagmentCrowdwork	-0.26
SmartphoneGivenCompletionCrowdwork	-0.23
Owned device type Speaker	0.17
IdealWorldSmartphoneCompleteCrowdwork	-0.17
IdealWorldSmartphoneManageCrowdwork	-0.17
No owned phone	0.16
No challenges desktop	0.15
SmartphoneCompleteCrowdwork	-0.14
No owned speaker	-0.14
SmartphoneUsagePlaceImportance	-0.14
SmartphoneManageCrowdwork	-0.14
UsageSmartspeakeroutsideofCrowdwork	0.13
SmartspeakerUsagePlaceImportance	0.13
SmartphoneUsageTimeImportance	-0.13
SmartspeakerUsageTimeImportance	0.12
UsageSmartphoneoutsideofCrowdwork	-0.12
Owned device type Phone	-0.12
PlatformsCompletingCrowdworkNon-DesktopDevices	-0.11
DesktopCompleteCrowdwork	0.11
PlatformsManagingCrowdworkNon-DesktopDevices	-0.11
UsageDesktopoutsideofCrowdwork	0.11
No challenges phone	-0.11
Difficult to manage work on speaker	0.10
No owned desktop	-0.10
Owned device type Desktop	0.10
Lack of tools on speaker	0.10
SmartspeakerManageCrowdwork	0.10
IdealWorldDesktopCompleteCrowdwork	0.10

Table 6.7: Top features of PCA Component 4 along with their loading factor within threshold absolute value 0.05

Features	**Loading Factor**
Owned device type Watch	0.25
No owned watch	-0.22
UsageSmartwatchoutsideofCrowdwork	0.21
SmartwatchUsagePlaceImportance	0.18
SmartwatchUsageTimeImportance	0.18
DesktopGivenManagmentCrowdwork	-0.17
Difficult to manage work on watch	0.16
Owned device type Speaker	-0.15
Lack of tools on watch	0.14
IdealWorldSmartwatchCompleteCrowdwork	0.14
IdealWorldSmartwatchManageCrowdwork	0.14
IdealWorldDesktopManageCrowdwork	-0.13
SmartphoneGivenManagmentCrowdwork	-0.13
SmartwatchManageCrowdwork	0.13
SmartwatchCompleteCrowdwork	0.13
Prohibits use of watch	0.13
SmartspeakerUsagePlaceImportance	-0.12
UsageSmartspeakeroutsideofCrowdwork	-0.12
SmartspeakerUsageTimeImportance	-0.12
DesktopGivenCompletionCrowdwork	-0.11
No owned speaker	0.11
DesktopManageCrowdwork	-0.11
IdealWorldDesktopCompleteCrowdwork	-0.11
Difficult to manage work on phone	-0.10
Lack of tools on phone	-0.10
DesktopCompleteCrowdwork	-0.10
UsageDesktopoutsideofCrowdwork	-0.10
IdealWorldSmartspeakerCompleteCrowdwork	-0.10
No owned desktop	0.10
Owned device type Desktop	-0.10

Table 6.8: Top features of PCA Component 5 along with their loading factor within threshold absolute value 0.05

Features	**Loading Factor**
Owned device type Speaker	0.27
No owned speaker	-0.22
UsageSmartspeakeroutsideofCrowdwork	0.21
SmartspeakerUsagePlaceImportance	0.20
SmartspeakerUsageTimeImportance	0.20
DesktopGivenManagmentCrowdwork	-0.18
IdealWorldSmartspeakerCompleteCrowdwork	0.16
IdealWorldDesktopManageCrowdwork	-0.16
SmartspeakerManageCrowdwork	0.15
IdealWorldSmartspeakerManageCrowdwork	0.15
DesktopCompleteCrowdwork	-0.15
IdealWorldDesktopCompleteCrowdwork	-0.15
DesktopGivenCompletionCrowdwork	-0.15
DesktopManageCrowdwork	-0.15
Difficult to manage work on speaker	0.14
SmartspeakerCompleteCrowdwork	0.14
UsageDesktopoutsideofCrowdwork	-0.14
Lack of tools on speaker	0.14
DesktopUsagePlaceImportance	-0.13
No owned desktop	0.13
Owned device type Desktop	-0.13
No challenges desktop	-0.12
No challenges phone	0.12
DesktopUsageTimeImportance	-0.12
Prohibits use of speaker	0.11
Difficult to manage work on phone	-0.11
Prohibits use of phone	-0.11
SpeakerGivenCompletionCrowdwork	0.10

smartwatches are used outside crowdwork contexts . It also mentions that people face significant limitations such as a lack of tools and explicit prohibitions in crowdwork settings, making them less effective for professional use. Similarly, smart speakers, while used in crowdwork encounter difficulties due to inadequate tool support and operational challenges in managing tasks effectively. It also reflects a notable interest in using smartphones predominantly for management such as finding and accepting tasks, communication as well as for reciving notifications rather than various completion tasks in ideal world , with a lesser but still significant interest in using smartwatches and smart speakers for crowdwork, contingent on the availability of better tooling.

The PCA component analysis reflects that tablets emerge as a versatile and well-integrated tool in workers' routines, while smartwatches and smart speakers, despite being commonly owned, face significant challenges due to inadequate tool support and operational limitations. Smartphones are predominantly used for management tasks, reflecting a nuanced preference and usage pattern among crowdworkers.

PCA Component 2 – High Focus on Smartwatch, Reluctance for Tablet due to Non-ownership

This PCA component analysis categorizes device utilization among crowdworkers into three distinct buckets based on their loading score positive value, highlighting preferences, challenges, and future potential for various devices used in crowdwork as shown in Table 6.5.

The first bucket consisting of features has the absolute value of PCA loading score from 0.18-0.29. This bucket focuses on smartwatches, indicating ownership and frequent use outside of crowdwork. The challenges noted include difficulties in managing tasks on the device, a lack of supporting tools, and restrictions against smartwatch use in crowdwork settings. Despite these issues, there is a strong interest in using smartwatches for both managing and completing tasks, provided that future improvements in tooling are made. Contextual factors like place and time are crucial in optimizing smartwatch usage in crowdwork scenarios.

The second bucket contains all the features with the absolute value of PCA loading score from 0.10-0.17. Features in this bucket reveal that smartwatches are actively used to manage and complete crowdwork tasks. Additionally, it discusses the use of speakers, primarily outside of crowdwork. While this device currently used for crowdwork management, it reflects challenges similar to those of smartwatches, such as a lack of tools and challenges in managing tasks. This bucket also underscores a high interest in using smartwatches, smartphones, and speakers for crowdwork management and completion, anticipating better tooling in the future. The importance of place and time remains significant in effectively utilizing speakers for crowdwork.

This component shows negative loading score for owning device type tablet and also using this currently along with using this device type for completion and mangament even in ideal world where high quality tooling is available for tablets.

The PCA component analysis highlights that smartwatches are not only used outside of crowdwork but also currently used in crowdwork. They face significant challenges within the crowdwork context due to inadequate tool support and operational limitations. Similarly, speakers, while used for crowdwork management, encounter similar difficulties. Despite these challenges, there is strong interest in using these devices for both managing and completing tasks, contingent on future improvements in high quality tooling in ideal crowdworld.

PCA Component 3 – Reluctance for Smartphone due to Non-ownership and Current and Desired use of Desktop for Completion

This PCA component analysis segments device usage among crowdworkers into three distinct buckets, each characterized by their unique set of devices, usage contexts, and associated challenges, as detailed in Table 6.6.

The first bucket, with PCA loading scores ranging from 0.17 to 0.26, highlights the ownership of speakers and the non-ownership of smartphones among crowdworkers. Consequently, smartphones are not used outside of crowdwork, within crowdwork, or in an ideal setup.

The second bucket with PCA loading score ranging from 0.10-0.16 notes that speakers are often used outside of crowdwork, with contextual factors like place and time playing significant roles. Additionally, using desktops for crowdwork does not currently present any challenges, indicating a stable integration of desktops in crowdwork environments. It also reflects high interest in using desktops for completing tasks in an ideal world where high-quality tooling is available. This bucket also discusses the challenges faced when using speakers for currently managing tasks, including difficulties in effective task management and a lack of appropriate tools in crowdwork settings. It also shows less enthusiasm for platforms supporting non-desktop devices for the completion and management of crowdwork.

This PCA component reveals ownership of speakers and the non-ownership of smartphones, leading to the non-use of smartphones across different contexts. It also reflects the prevalent and stable use of desktops for crowdwork, highlighting high interest in desktops under ideal world, while noting the challenges associated with using speakers for current use in crowdwork setting.

PCA Component 4 – High Focus on Smartwatch, Reluctance for Speaker and Desktop due to Non-ownership

This PCA component analysis dissects device usage among crowdworkers into three distinct buckets, outlining how various devices are utilized and the specific challenges encountered, as presented in Table 6.7.

The first bucket consisting of features has the absolute value of PCA loading score from 0.17-0.25. This segment primarily focuses on the use of smartwatches. It notes that smartwatches are predominantly used outside of crowdwork. Contextual factors like location and time play significant roles when this device type is used for completion and management of crowdwork. It also reflects less interest in using desktop for given varities of managment crowdwork including finding and accepting tasks, communication with requesters and other workers as well as receiving notification.

The second bucket consisting of features has the positive value of PCA loading score from 0.10-0.16. The second bucket mentions the current use of this device type for crowdwork and also highlights the challenges faced while using smartwatch for completing and managing crowdwork currently. They include difficulty in managing tasks on such a small interface, lack of appropriate tools, and regulatory restrictions against their use in certain crowdwork activities. Despite these issues, there is a notable interest in using smartwatches more extensively for both managing and completing tasks, especially if high quality tooling become available, reflecting a positive view of their potential in an ideal scenario with better support.

This bucket also reflects a lack of ownership of desktops and smart speakers. Consequently, these devices are not used outside of crowdwork, within crowdwork, or in an ideal setup. There is less enthusiasm for using smartphones for various management tasks, including finding and accepting tasks, communication with requesters and other workers, and receiving notifications. Additionally, there is less emphasis on the challenges of using smartphones currently for completing and managing crowdwork.

The PCA component analysis reveals that smartwatches are used significantly outside of crowdwork. Despite current challenges like small interfaces, lack of tools, and regulatory restrictions, smartwatches have high interest for crowdwork. Additionally, the analysis highlights the absence of desktops and smart speakers.

PCA Component 5 – High Focus on Smart Speaker with Reluctance for Desktop due to Non-ownership

This PCA component analysis categorizes device usage among crowdworkers into three buckets, highlighting how various devices are employed and the particular challenges faced, as detailed in Table 6.8.

he first bucket, with PCA loading scores from 0.15-0.27, focuses on smart speakers. It notes their frequent use outside of crowdwork, emphasizing the importance of contextual factors like place and time. There is a high interest in using smart speakers for managing and completing crowdwork with high-quality tooling. However, there is less interest in using desktops for managing and completing various crowdwork tasks, including finding and accepting tasks, communication, and various types of completion tasks like surveys and testing even in ideal world where high quality tooling avilable for desktop.

The second bucket, with PCA loading scores from 0.10-0.14, highlights challenges in using speakers currently for crowdwork, including difficulties in task management and a scarcity of tailored tools. It also notes the lack of desktop ownership, resulting in no use for crowdwork and even outside of crowdwork. In contrary, smartphones are currently used for crowdwork without significant issues.

The analysis reveals that smart speakers are frequently used outside of crowdwork and are of high interest for future use in crowdwork, provided high quality tools are available in ideal world. Desktops are not favored for managing and completing crowdwork tasks, due to non ownership of this device.

6.5.4 Description of Clusters of Prolific Crowdworkers

Here, we describe the six groups of workers based on their preference of using devices on different stages of work as the following based on PCA mean and median values.

Cluster 1 – Desktop Enthusiasts with Aspirations for Smartphone

This cluster comprises 273 workers, as presented in Tables 6.9 and 6.10 show that PCA1 emerges with the highest absolute magnitude (mean = 1.58, median = 1.72), followed by PCA4 (mean = 0.80, median = 0.78), both oriented in negative directions. This is succeeded by PCA3, PCA2, and PCA5, where PCA3 and PCA5 also lean negative, while PCA2 takes a positive direction. The negative orientation of PCA1 reveals that most workers in this cluster lack ownership of several non-desktop devices, such as tablets, smartwatches, and smart speakers. This lack of ownership correlates with a minimal use of these devices for crowdwork. Additionally, even in an ideal scenario where high-quality tools are available, these workers show little enthusiasm for using smartphones for their tasks.

PCA4's negative direction uncovers a nuanced dynamic. A significant portion of these workers show high interest in completion and management of crowdwork, such as transcription (see Figure 6.26) along with their current use. They also express a willingness to integrate smartphone mostly for management into their workflow in ideal world. However, there remains a pronounced disinterest in smartwatches, mainly driven by non-ownership, highlighting a gap between available devices and user preferences.

The negative orientation of PCA3 underscores the current use of smartphones for crowdwork and a high interest in using them for crowdwork management and completion. Yet, this cluster also includes members who do not own smart speakers and desktops.

In contrast, PCA2's positive direction signals current and desired use of smartwatches among some workers, despite the lack of tablet ownership within the group.

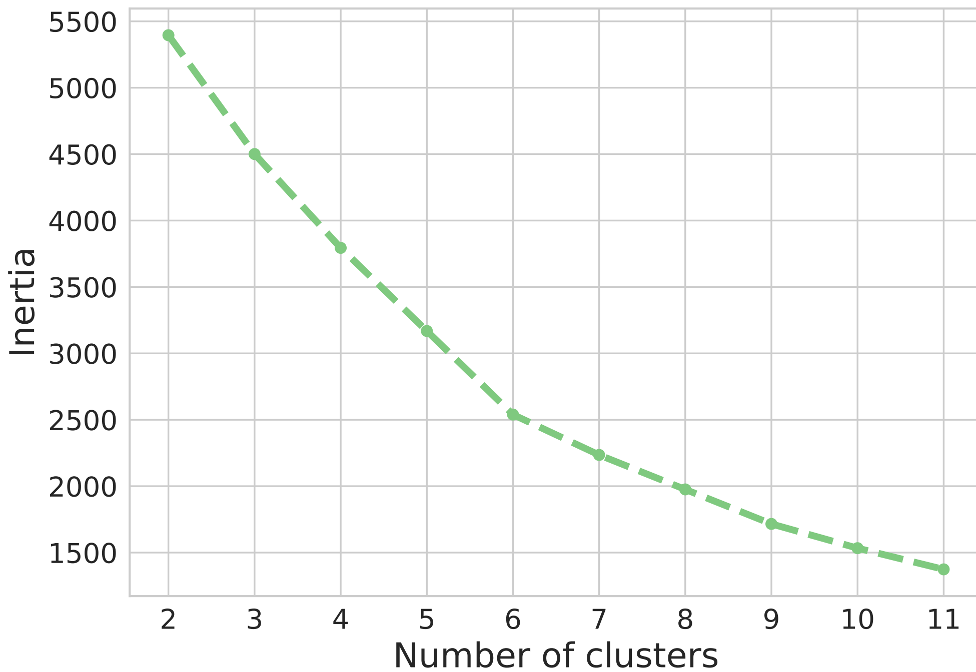


Figure 6.23: Elbow plot which shows the optimized number of clusters are 6 for our dataset

Table 6.9: PCA Mean Values along with six clusters and their size

PCA1	PCA2	PCA3	PCA4	PCA5	Cluster	Cluster Size
-1.58	0.31	-0.46	-0.80	-0.17	1	273
1.40	-1.30	-0.17	-0.62	-0.06	2	239
2.42	0.60	0.23	0.67	-0.06	3	198
-0.53	2.13	0.13	0.60	-0.31	4	127
-2.05	-0.98	1.94	0.51	-0.15	5	110
-1.56	-0.97	-1.92	1.76	2.30	6	57

Table 6.10: PCA Median Values along with six clusters and their size

PCA1	PCA2	PCA3	PCA4	PCA5	Cluster	Cluster Size
-1.72	0.24	-0.59	-0.78	-0.47	1	273
1.44	-1.46	-0.37	-0.62	-0.46	2	239
2.44	0.42	0.25	0.62	0.188	3	198
-0.62	2.03	0.00	0.70	-0.59	4	127
-2.59	-0.99	2.16	0.37	-0.41	5	110
-2.31	-0.60	2.29	1.55	2.31	6	57

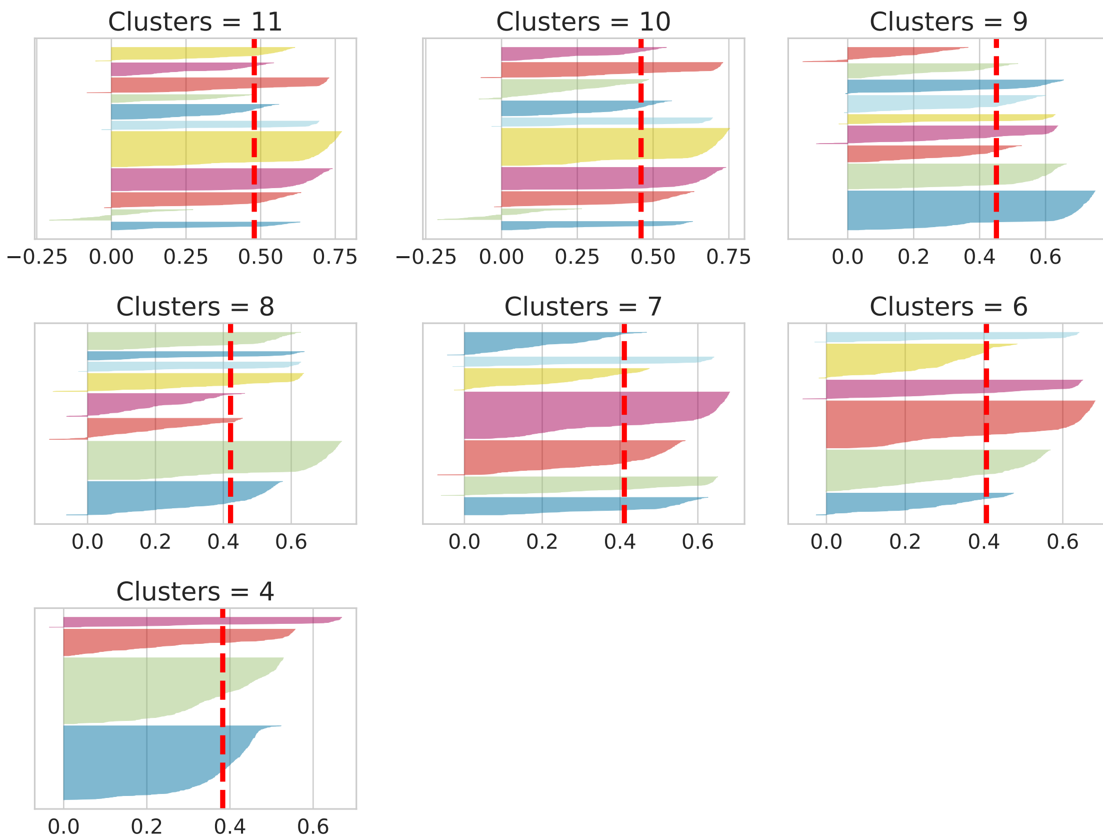


Figure 6.24: Siloutee score plot which shows the optimized number of clustering is 6 in our case

Lastly, PCA5's negative direction highlights the current use and interest in desktops in ideal world, while also pointing out a notable disinterest in smart speakers, primarily due to non-ownership.

This cluster of 273 workers is characterized by their strong attachment to desktops and a lack of interest in several other device types, largely due to non-ownership. Despite this, there is a notable openness to integrating smartphones into their workflow in an ideal world. However, the disinterest in smartwatches and other non-desktop devices underscores a gap between device availability and user preferences, pointing to areas where device ownership could potentially enhance their productivity and satisfaction.

Cluster 2 – Diverse Device Users with a Preference for Tablet and Desktop

This cluster comprises 239 workers, as presented in Tables 6.9 and 6.10.

With the highest absolute magnitude according to the mean (mean = 1.40, median = 1.44), PCA1 stands out with a positive direction. This indicates a broad utilization of various devices, with a strong focus on tablets. These workers show a keen interest in using smartphones primarily for management tasks and smartwatches for task completion in an ideal world.

PCA2 follows closely, boasting the highest absolute magnitude according to the median (median = 1.46, mean = 1.30), but it takes a negative direction. This suggests that these workers not only currently use tablets extensively but also have a high preference for using them for both completing and managing crowdwork. However, some members of this group do not own smartwatches, reflecting a reluctance to use this device type.

In the next sequence, PCA3, PCA4, and PCA5 (with negative directions) show intriguing patterns. PCA3 highlights the current and desired use of smartphones for both completion and management, but also indicates a lack of ownership of smart speakers and desktops among these workers. PCA4, with a significant negative direction, underscores the current use and strong interest in using desktops, mostly for management than completion. This preference stands in stark contrast to the group's disinterest in smartwatches, attributed to non-ownership. Finally, PCA5, also negative, focuses on the current use and interest in desktops. It points out a notable disinterest in smart speakers, primarily due to non-ownership.

This cluster of 239 workers is characterized by diverse device usage patterns with a strong preference for tablets and desktops. Workers exhibit broad utilization of various devices, especially tablets, for managing and completing tasks. Despite a keen interest in using smartphones and smartwatches in ideal scenarios, non-ownership of smartwatches limits their use. Desktops are crucial for task management, but smart speakers are less favored, primarily due to non-ownership. These preferences and gaps in device ownership significantly influence how these workers approach their digital tasks.

Cluster 3 – Tablet Enthusiasts and Smartwatch Aficionados

In the evolving landscape of digital device usage, this cluster of 198 workers stands out with distinct preferences and usage patterns. Detailed in Tables 6.9 and 6.10, these workers' device inclinations are illuminated through Principal Component Analysis (PCA). PCA1, with the highest absolute magnitude and positive direction (mean = 2.42, median = 2.44), reveals a strong current usage of tablets, accompanied by a high preference for this device type in an ideal world (see Figure 6.25. These workers also exhibit a keen interest in using smartphones mainly for management tasks and smartwatches for task completion in an ideal scenario.

PCA2, in a positive direction, highlights a strong interest using smartwatches in ideal world for both managing and completing crowdwork mostly rather than their current use of this device in crowdwork. However, it also notes a gap in tablet ownership among some workers.

PCA3, also positive, indicates a significant reliance on desktops for crowdwork, primarily for task completion both currently and in an ideal world along with current use of smart speaker for managing crowdwork. This reliance is paired with a reluctance to use smartphones, primarily due to non-ownership. PCA4 shows a strong focus on smartwatches, with high interest and current usage of workers, especially in an ideal world. Conversely, there is a noted reluctance to use desktops and smart speakers, again primarily due to non-ownership. Finally, PCA5, with a positive direction according to the median, reveals a high level of interest in using smart speakers, driven by ownership among the workers. However, this component also points out the non-ownership of desktops, which restrics their use of this device type in crowdwork.

This cluster of 198 workers showcases a dynamic and varied engagement with digital devices. Tablets emerge as a favorite, with high current usage and preference in an ideal world. Workers also show robust engagement on smartwatches. Desktops are crucial for task completion, though their usage is limited by non-ownership among some workers. Smartphones are favored for management tasks but face reluctance due to non-ownership. Lastly, smart speakers hold interest primarily among those who already own them, highlighting how device ownership directly influences preferences and usage patterns in crowdwork.

Cluster 4 – Smartwatch Enthusiasts with Current and Desired use of Desktop who own them

This cluster of 127 workers stands out with unique inclinations and usage patterns. Detailed in Tables 6.9 and 6.10, their preferences are illuminated through Principal Component Analysis (PCA).

PCA Component 2, with the highest absolute magnitude (mean = 2.13, median = 2.03), highlights a high interest in using smartwatches for both completing and managing crowdwork in an ideal world. Despite their enthusiasm, these workers face challenges due to a lack of appropriate tools, as shown in Figure 6.25. They also emphasize the importance of contextual factors, such as the place of work, when using smartwatches for crowdwork (Figure 6.26). Additionally, there is a notable lack of tablet ownership within this group, resulting in no use of tablets for their tasks.

PCA4, with a positive direction (mean = 0.60, median = 0.70), further underscores a strong affinity for smartwatches, indicating significant usage. This component contrasts sharply with a marked reluctance to use desktops and smart speakers, primarily due to the absence of these devices among the workers.

PCA1, in a negative direction, reflects a significant deficiency in the ownership of various devices, including tablets, smartwatches, and smart speakers. The negative values suggest no use of these technologies in crowdwork, highlighting how non-ownership limits their digital engagement.

PCA5, also in a negative direction, focuses on the current use and interest in desktops. However, it points out a notable disinterest in smart speakers, which is primarily linked to non-ownership. This lack of smart speaker usage underscores how device availability influences their preferences.

This cluster of 127 workers reveals a complex relationship with their digital devices. Their strong interest in smartwatches for managing and completing crowdwork is evident, despite facing challenges due to insufficient tools and the importance of contextual factors. The lack of tablet ownership further restricts their device usage. While there is significant smartwatch usage, desktops and smart speakers see minimal interest due to non-ownership. This cluster's story highlights how the availability and ownership of devices play a crucial role in shaping their preferences and usage patterns in the realm of crowdwork.

Cluster 5 – Desktop-Dependents with Selective Tablet and Smartwatch Preferences

This cluster of 110 workers, as detailed in Tables 6.9 and 6.10 presents a nuanced and complex pattern of device usage and preferences.

PCA1, with the highest magnitude (mean = 2.05, median = 2.58) and a negative direction, highlights a significant deficiency in the ownership of devices such as tablets, smartwatches, and smart speakers. This lack of ownership results in minimal use of these technologies for crowdwork, underscoring how non-ownership limits their digital engagement.

PCA3 follows closely with a high magnitude (mean = 1.94, median = 2.16) and a positive direction. This component indicates a reliance on desktops for crowdwork, primarily for task completion both currently and in an ideal world. However, this reliance is coupled with a reluctance to use smartphones, again due to non-ownership, as shown in Figure 6.25.

PCA2, with a negative direction, reveals that some workers currently use tablets and have a high preference for using them for both completing and managing crowdwork in an ideal world. However, the reluctance to use smartwatches persists due to non-ownership among some group members.

PCA5, also in a negative direction, focuses on the current use and interest in desktops. It highlights a notable disinterest in smart speakers, primarily linked to non-ownership, indicating how device availability influences their preferences and usage patterns.

PCA4, with a positive direction, showcases a strong affinity for smartwatches, indicating significant usage among those who own them. This positive interest in smartwatches contrasts with a marked reluctance to use desktops and smart speakers, again attributed to the absence of these devices among the workers.

This cluster of 110 workers presents a complex and nuanced pattern of device usage and preferences. While there is a significant reliance on desktops for task completion, the lack of ownership of tablets, smartwatches, and smart speakers limits the use of these devices for crowdwork. Tablets are favored by those who own them, especially for managing and completing tasks in an ideal world. However, smartwatches are primarily used by their owners, and there is a clear reluctance to use desktops and smart speakers due to non-ownership. This cluster's story highlights the critical role of device ownership in shaping digital engagement and preferences in the context of crowdwork.

Cluster 6 – Smart Speaker Enthusiasts with Smartphone and Smartwatch Reliance

This cluster, comprising 57 workers as referenced in Tables 6.9 and 6.10.

PCA Component 5 stands out with the highest absolute magnitude (mean = 2.30, median = 2.31) and a positive direction. This indicates a high level of interest in using smart speakers, driven by ownership among these workers. However, this component also highlights the non-ownership of desktops, which restricts the use of desktops for crowdwork tasks, as shown in Figure 6.25.

PCA1, with a negative direction, reflects a significant deficiency in the ownership of various devices, including tablets, smartwatches, and smart speakers. The negative values suggest minimal use of these technologies in crowdwork due to non-ownership.

PCA3, also in a negative direction, underscores the current use of smartphone in crowdwork and the desired use of these devices for both task completion and management. However, it also reflects a lack of ownership of smart speakers and desktops, limiting their usage.

PCA4, with a positive direction, indicates a strong affinity for smartwatches, showcasing significant usage among those who own them. This component contrasts with a marked reluctance to use desktops and smart speakers, again attributed to non-ownership.

PCA2, in a negative direction, reveals that while some workers currently use tablets and have a high preference for using them for both completing and managing crowdwork in an ideal world, there is also a reluctance to use smartwatches due to non-ownership among some group members.

This cluster of 57 workers illustrates that smart speakers are highly favored by those who own them, while the non-ownership of desktops restricts their use in crowdwork. A significant deficiency in the ownership of tablets, smartwatches, and smart speakers results in minimal usage of these devices. Despite this, there is a notable reliance on currently owned devices for both task completion and management. Smartwatches see significant use among their owners, but desktops and smart speakers face reluctance due to non-ownership. Smartphones are essential for crowdwork tasks, yet the absence of smart speakers and desktops among many workers limits their use. Tablets are used and preferred by some for crowdwork tasks, yet the absence of smartwatches among many workers limits their use.

The analysis of the six clusters reveals unique preferences and behaviors in their use of technological tools for managing and completing crowdwork. Cluster 1 exhibits the least interest in tablets, with very few tablet owners, but shares with Cluster 5 a strong reliance on desktops. Cluster 2 stands out with a strong preference for tablets and less smartwatch ownership, placing less importance on contextual factors like location when using smartwatches. Cluster 3 uses smartwatches and tablets but relies more heavily on tablets for their tasks, facing challenges with insufficient smartwatch tools. Cluster 4, while also using smartwatches and tablets, shows a moderate preference for tablets and similar challenges with smartwatch tools. Cluster 5 is notable for having fewer smartwatch owners and less enthusiasm for using them, indicating a more varied approach to device engagement. Cluster 6 is characterized by a strong preference for smart speaker usage and an openness to smartphone and smartwatch and lack of ownership of desktop. These distinctions highlight the diverse levels of technology adoption and reliance among the clusters.

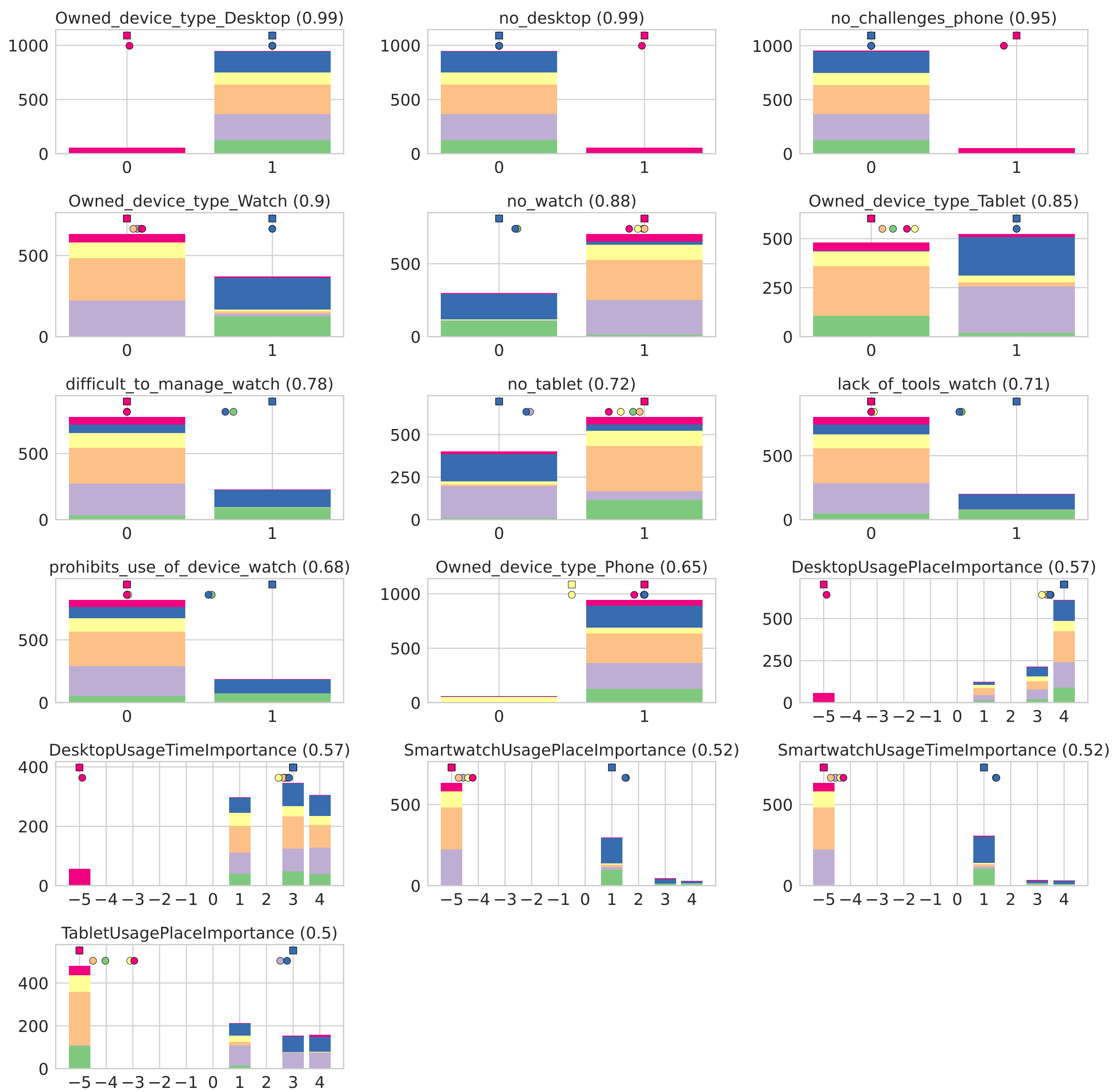


Figure 6.25: Visualization of all the significant features as shown in Table 6.11 within threshold effect size of more than 0.5 across all six clusters of crowdworkers

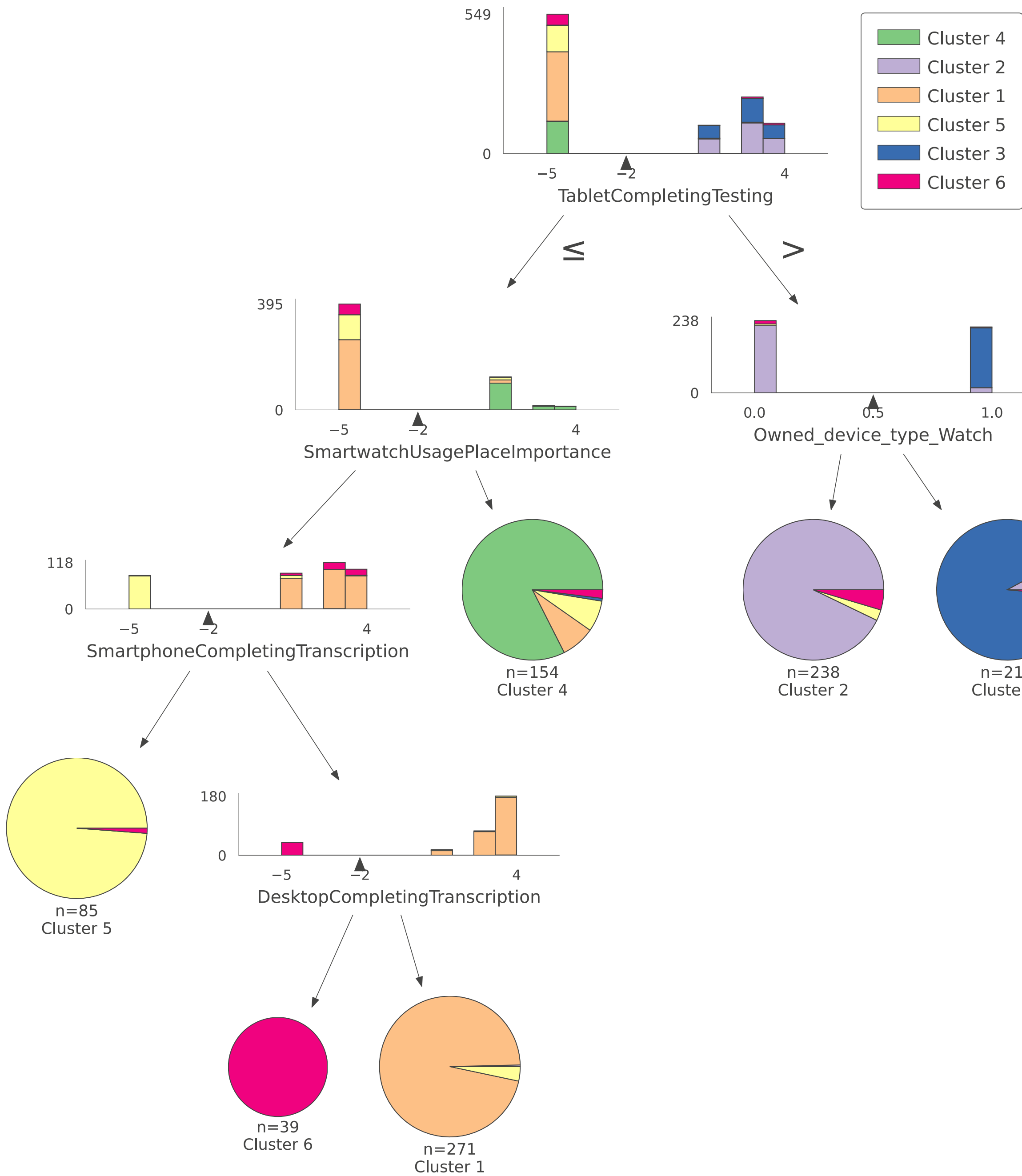


Figure 6.26: Decision Tree which showsall unrolled features that differentiate clusters of crowdworkers

6.6 Comparison of Clusters of Crowdworkers between MTurk and Prolific

In this analysis, we explore the similarities and differences between clusters of crowdworkers on Amazon Mechanical Turk (MTurk) and Prolific, focusing on their device preferences and activities across various stages of crowdwork.

MTurk features three main clusters: the traditionalists who primarily use desktops or workstation but have limited openness to incorporate smartphones and tablets in their crowdwork practices; the second largest group, which uses mainly traditional devices such as desktop or workstation, however expresses a need for enhanced management capabilities on smartphone and tablet and the non-workstation enthusiasts, who favor a broader integration of non-workstation devices and express a high interest in managing tasks across all devices.

Prolific hosts six diverse clusters. The largest cluster, Desktop Enthusiasts, is similar to MTurk's Traditionalists, primarily using desktops but also showing openness to using smartphones. Additionally, Cluster 5, the Desktop dependents, shows a strong dependence on desktops or workstations with selective interest in tablets and smartwatches.

The second largest cluster in Prolific, Diverse Device Users, utilizes a variety of devices, with a preference for tablets and desktops. The fourth largest cluster in Prolific are interested in using smartwatches with current and desired use of Desktop who own them.

Prolific also includes three groups who prefer using non-desktop devices as shown in MTurk third cluster. Tablet Enthusiasts who primarily uses tablets and shows a high preference for using smartwatches in an ideal scenario and Cluster 6, who are smart speaker enthusiasts with reliance on smartphones and smartwatches.

However, there are also distinct differences between the clusters on these platforms. MTurk has three distinct clusters, whereas Prolific has six, suggesting a more diverse range of device preferences among Prolific workers. Prolific's clusters exhibit more specific preferences, such as the Tablet Enthusiasts , Smart Watch Enthusiasts, which are not as prominently observed in MTurk's clusters. While MTurk's clusters primarily focus on traditional and non-traditional device usage mostly on smartphone and tablet, Prolific's clusters show a more nuanced combination of device preferences, including specific interests in smartwatches and smart speakers.

In summary, while both platforms host workers with traditional and non-traditional device usage preferences, Prolific's clusters display a greater diversity and specificity in device preferences, reflecting a broader spectrum of crowdwork practices.

6.7 Correlation of Diferent types of Demographics and Crowdworkers' Work Practices

In this section we discuss if there exists any significant correlations between crowdworkers' device usage and general demographics, technical factors , financial factors as well as profession outside of crowdwork.

6.7.1 Method

In this analysis, we excluded data from participants who revoked consent or data got expired for any demographic column that is collected by Prolific to enhance the reliability and validity of our correlation findings. After this data cleansing process, we retained data from 755 participants out of the original 1004.

To test our hypothesis, we employed a variety of statistical methods tailored to the nature of our data. For categorical variables, we utilized chi-square tests to assess the corelation . For interval and ratio variables, we applied Pearson correlation coefficients to measure the strength and direction of linear relationships. Additionally, we used ANOVA to compare means across multiple groups when data type is categorical and interval ratio data when it is normally distributed and there is no variance by doing Normality test and Levene's test for homogeneity of variances, otherwise we used Kruskal-Wallis Test. Missing data due to non-ownership of specific devices were accounted for in our analyses to ensure robust and reliable results. In this case we sort the features first by first adjusted p-value, then by corrected effect size and then by statistic (which is corelation) in order to first identify associations that are statistically significant, then prioritize them based on their practical significance (corrected effect size), and finally examine the strength of the linear relationships (correlation). It helps to ensure that we focus on the most meaningful and relevant associations between all type of demographics and work practices while considering both statistical and practical significance. In this case, we discuss the top 5 most significant correlated features.

6.7.2 Correlation of General Demographics and Crowdworkers Work Practices

In this analysis, we explore the correlation between general demographics and crowdworkers' device usage in their work practices, focusing on key findings. Table 6.12 highlights the top five most significant features that show correlations, including regional variations, current and ideal use of desktops, and the relationship between gender and smartphone usage for managing crowdwork.

Gender and Device Usage

Female crowdworkers use smartphones significantly more often to manage crowdwork than their male counterparts. This finding is supported by a chi-squared test ($\chi^2 = 38.51$, $p < 0.001$, Corrected Effect size = 0.21), indicating a notable gender difference in smartphone usage for task management as shown in Figure 6.27.

Regional Device Preferences

Across all regions, Canada and Africa exhibit the highest percentage of people currently using desktops very often to always for managing ($\chi^2 = 138.17$, $p < 0.001$, Corrected Effect size = 0.17) and completing crowdwork ($\chi^2 = 110.54$, $p < 0.001$, Corrected Effect size = 0.14). This trend is more pronounced for task completion than for task management in all regions. In contrast, the USA, Oceania, and the UK have the lowest percentage of people currently using desktops to complete and manage crowdwork as shown in Figure 6.28 and Figure 6.30.

In an ideal scenario with high-quality tooling available for desktops, the highest percentage of people in Africa prefer using this device type very often to always for managing crowdwork, slightly surpassing Americas and Canada. Other regions such as the Americas, Asia, and

Table 6.11: All the significant features along with their p-value within threshold effect size of more than 0.5 across all six clusters of crowdworkers

Feature	Display	p-value	Effect-size	Corrected-effect-size
Have the device Desktop	$\chi^2(5, 1004) = 985.3$	< 0.001	0.991	0.989
Don't have the device Desktop	$\chi^2(5, 1004) = 985.3$	< 0.001	0.991	0.989
No current challenges using Smartphone for Crowdwork	$\chi^2(5, 1004) = 911.1$	< 0.001	0.953	0.950
Have the device Smartwatch	$\chi^2(5, 1004) = 821.4$	< 0.001	0.905	0.902
Don't have the device Smartwatch	$\chi^2(5, 1004) = 786.1$	< 0.001	0.885	0.882
Have the device Tablet	$\chi^2(5, 1004) = 726.7$	< 0.001	0.851	0.848
Difficult to manage crowdwork on Smartwatch	$\chi^2(5, 1004) = 614.2$	< 0.001	0.782	0.779
Don't have the device Tablet	$\chi^2(5, 1004) = 529.5$	< 0.001	0.726	0.723
Lack of tools to work on Smartwatch	$\chi^2(5, 1004) = 510.0$	< 0.001	0.713	0.710
Not able to use Smartwatch for Crowdwork because of explicit prohibition	$\chi^2(5, 1004) = 469.7$	< 0.001	0.684	0.681
Have the device Phone	$\chi^2(5, 1004) = 430.7$	< 0.001	0.655	0.651
Importance of Place while using Desktop in Crowdwork	$\chi^2(15, 1004) = 1004.9$	< 0.001	0.578	0.574
Importance of Time while using Desktop in Crowdwork	$\chi^2(15, 1004) = 999.4$	< 0.001	0.576	0.573
Importance of Place while using Smartwatch in Crowdwork	$\chi^2(15, 1004) = 828.7$	< 0.001	0.525	0.521
Importance of Time while using Smartwatch in Crowdwork	$\chi^2(15, 1004) = 826.7$	< 0.001	0.524	0.520
Importance of Place while using Tablet in Crowdwork	$\chi^2(15, 1004) = 779.7$	< 0.001	0.509	0.505

Table 6.12: Significant features that show the correlation between various demographics, technical factors, and professions outside of crowdwork with crowdworkers' device usage for work practices

Category	Feature	Statistic	Raw Effect Size	Corrected Effect Size
Demographic Factors	(Gender, Use of Smartphone to currently manage crowdwork)	$\chi^2(5, 755) = 38.5, p < 0.001$	0.23	0.21
	(Region of crowdworkers, Use of Desktop to currently manage crowdwork)	$\chi^2(35, 755) = 138.2, p < 0.001$	0.19	0.17
	(Region of crowdworkers, Preference of Desktop to manage crowdwork)	$\chi^2(35, 755) = 124.5, p < 0.001$	0.18	0.15
	(Region of crowdworkers, Use of Desktop to currently complete crowdwork)	$\chi^2(35, 755) = 110.5, p < 0.001$	0.17	0.14
	(Region of crowdworkers, Use of Smartphone to currently complete crowdwork)	$\chi^2(35, 755) = 109.8, p < 0.001$	0.17	0.14
Technical Factors	(Use of Desktop outside crowdwork, Use of Desktop to currently complete crowdwork)	$\chi^2(16, 755) = 358.7, p < 0.001$	0.35	0.35
	(Use of Desktop outside crowdwork, Preference of Desktop to complete crowdwork)	$\chi^2(16, 755) = 233.7, p < 0.001$	0.29	0.28
	(Use of Desktop outside crowdwork, Use of Desktop to currently manage crowdwork)	$\chi^2(16, 755) = 132.5, p < 0.001$	0.22	0.20
	(Use of Desktop outside crowdwork, Preference of Desktop to manage crowdwork)	$\chi^2(16, 755) = 126.8, p < 0.001$	0.21	0.20
	(Prefer to use known technology, Preference of smart speaker to complete crowdwork)	$\chi^2(20, 755) = 114.4, p < 0.001$	0.19	0.18
Profession outside Crowdwork	(Profession of crowdworker, Preference of Smart speaker to manage crowdwork)	$\chi^2(40, 755) = 81.1, p = 0.005$	0.14	0.10

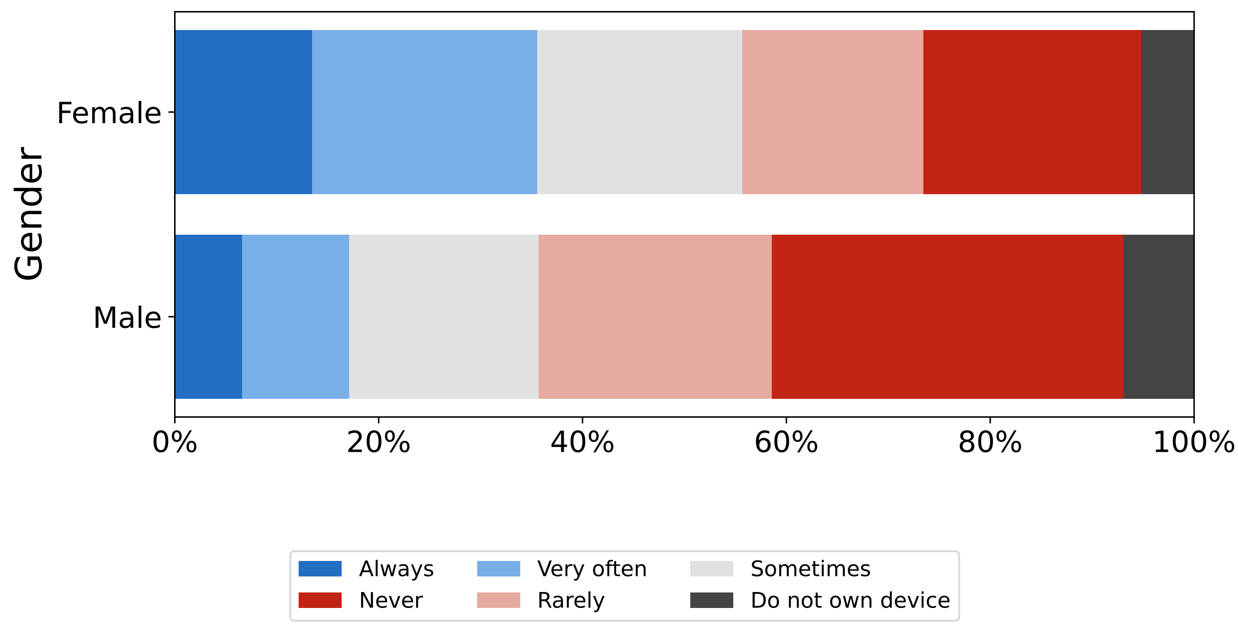


Figure 6.27: Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to gender and current smartphone use to manage crowdwork

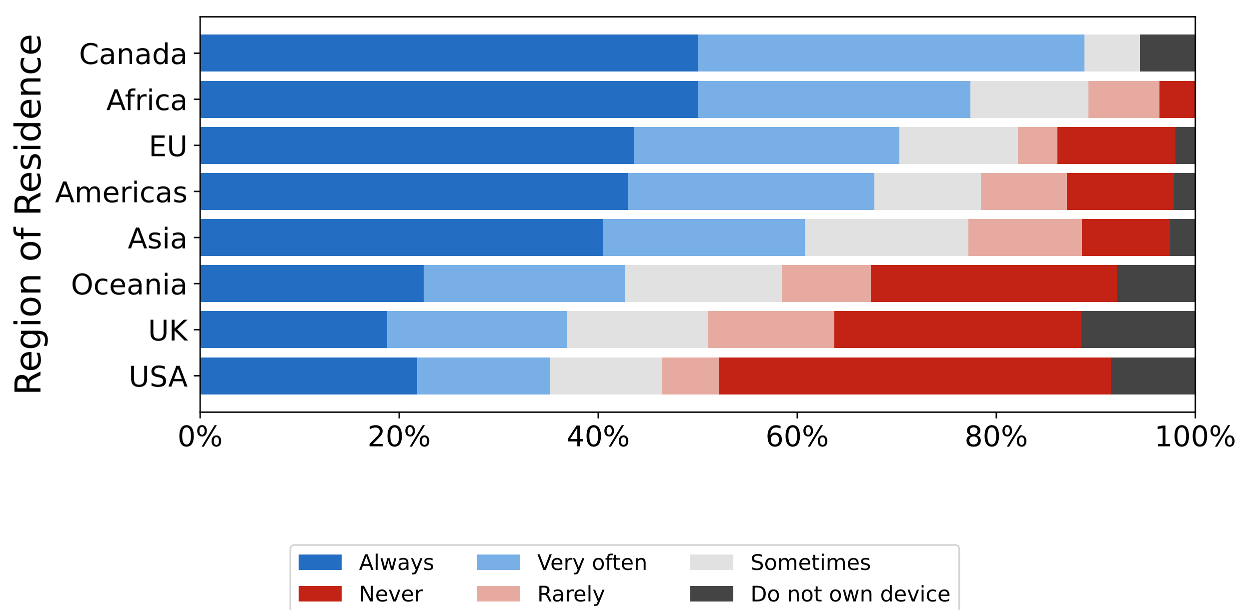


Figure 6.28: Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to currently manage crowdwork

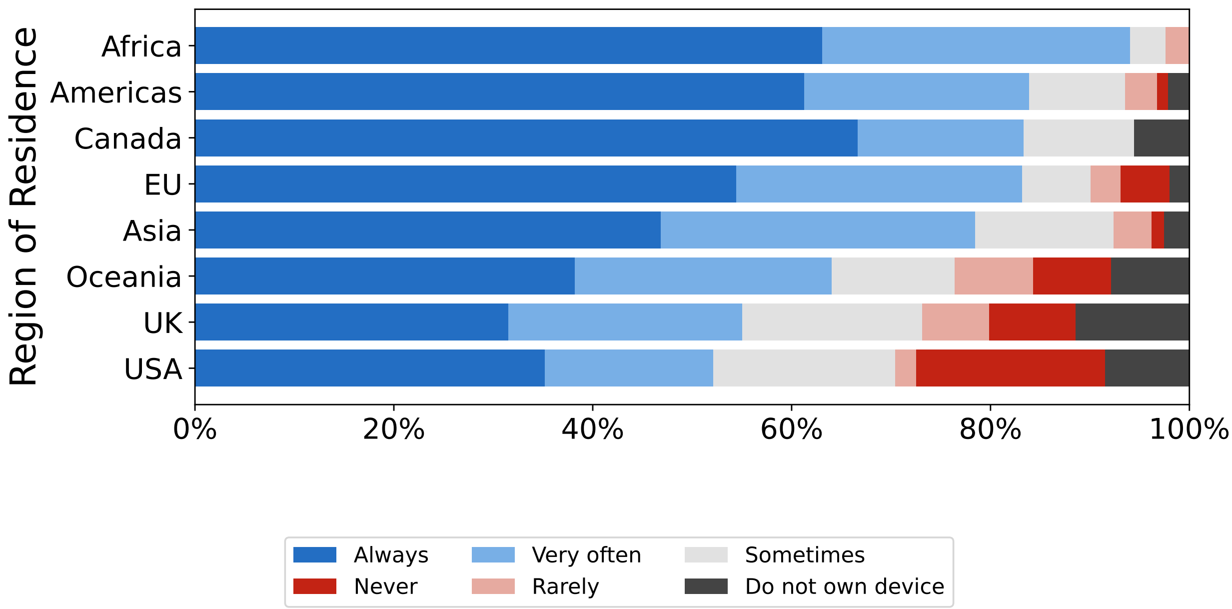


Figure 6.29: Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to manage crowdwork in ideal world

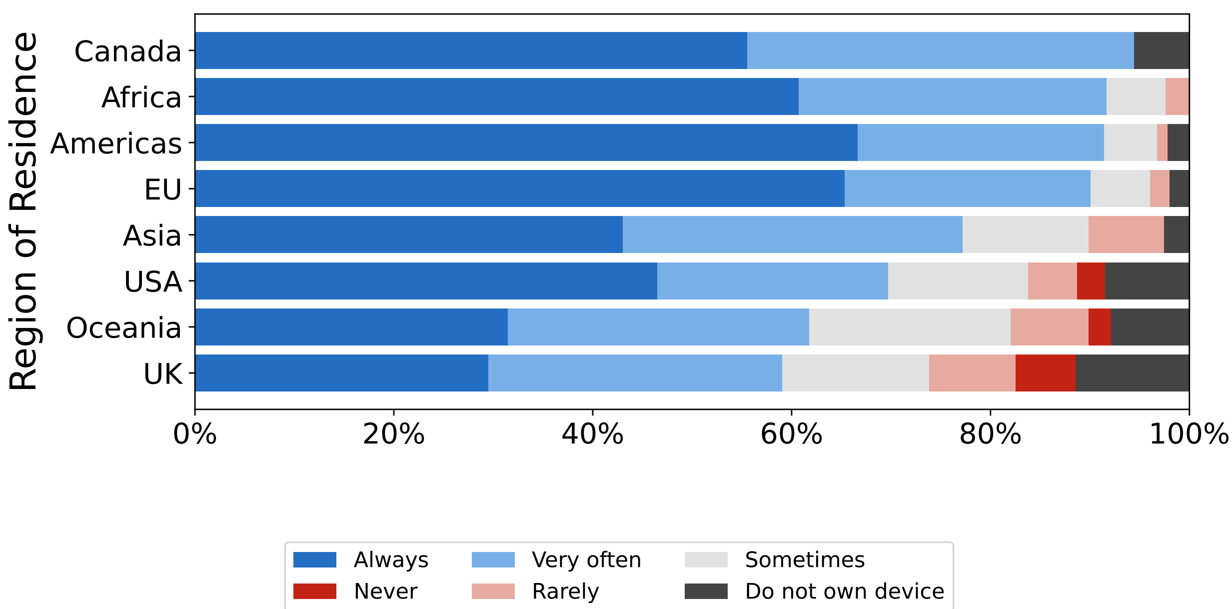


Figure 6.30: Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using desktop to complete crowdwork currently

Europe also show a higher preference for desktop usage compared to Oceania, the UK, and the USA. This trend is evident in managing crowdwork ($\chi^2 = 124.46$, $p < 0.001$, Corrected Effect size = 0.15) as shown in Figure 6.29.

Oceania and UK (see Figure 6.31) stand out with the highest percentage of people currently using smartphones very often to always for completing crowdwork. Additionally, regions such as the Africa, USA, Asia, and Americas also have around more than 20 percentage of the population using smartphones for this purpose. In contrast, Canada and Europe have the lowest percentage of workers using smartphones very often to always for current task completion ($\chi^2 = 109.75$, $p < 0.001$,Corrected Effect size = 0.14).

These correlations highlight the diverse device preferences and work practices across different demographics and regions among crowdworkers. The findings suggest significant regional differences in device usage for crowdwork, with Canada and Africa favoring desktops, and Oceania and Africa showing a strong preference for smartphones. The USA and UK, with the lowest desktop usage, indicate a reliance on non-desktop devices, reflecting broader trends in device ownership and accessibility.

6.7.3 Correlation of Technical Factors and Crowdworkers Work Practices

Our analysis revealed the top five most significant correlations between technical demographics and crowdworkers' device usage for managing and completing crowdwork (see Table 6.12). This include using desktop outside of crowdwork as well as for comleting and managing crowdwork along with technical interest and openness to use non-traditional device like speaker in crowdwork enviornment.

Desktop Usage outside of Crowdwork vs Desktop Usage for Completing and Managing Crowdwork

Workers who use desktops outside of crowdwork, regardless of frequency, tend to use desktops more for completing crowdwork than for managing it. This pattern remains consistent even in an ideal scenario where high-quality tooling is available for desktops. Workers who frequently

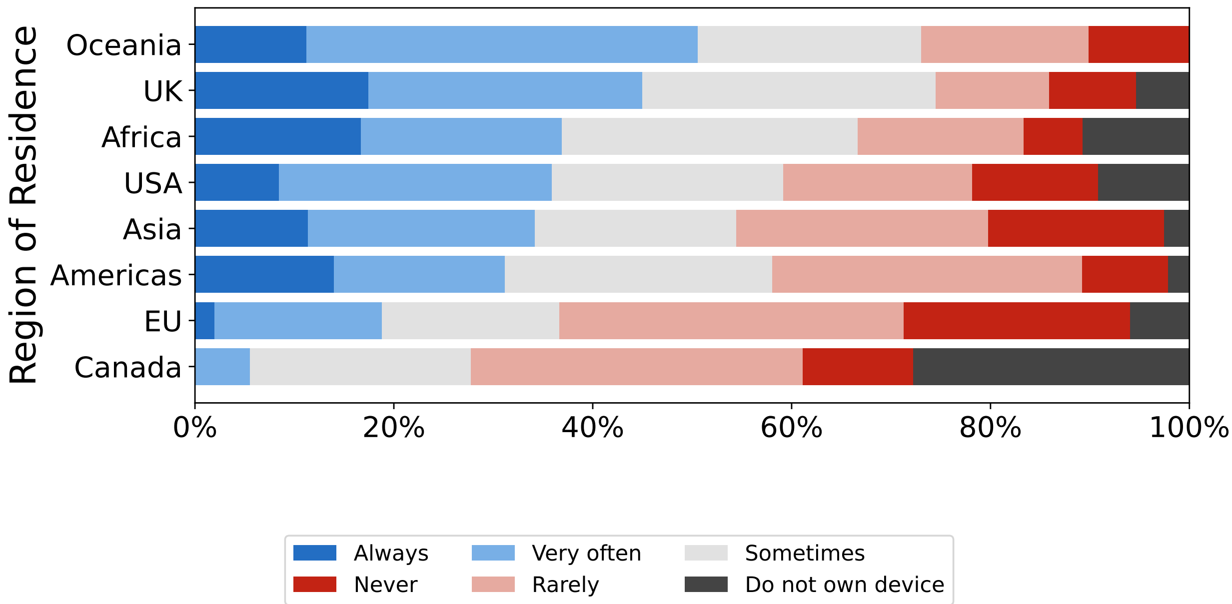


Figure 6.31: Visualization of significant feature as shown in Table 6.12 between general demographics and crowdworkers' work practices that have correlation related to region where crowdworkers live and frequency of using smartphone to complete crowdwork currently

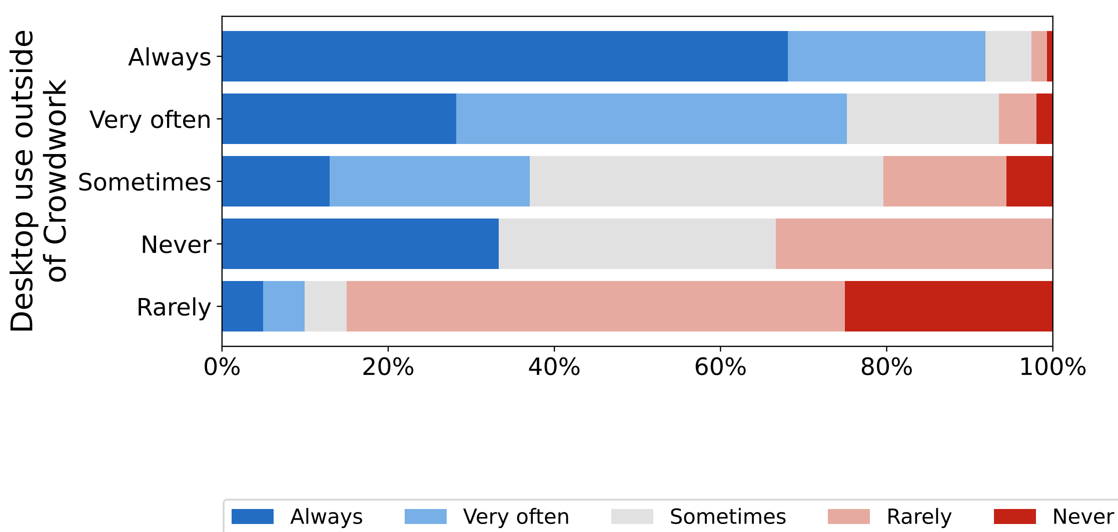


Figure 6.32: Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to currently complete crowdwork

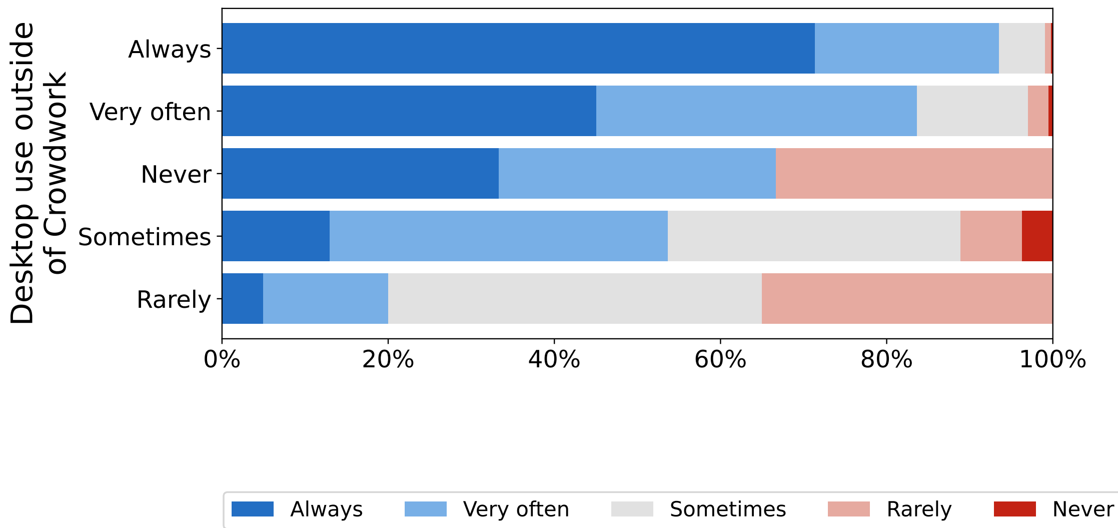


Figure 6.33: Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to complete crowdwork in ideal world

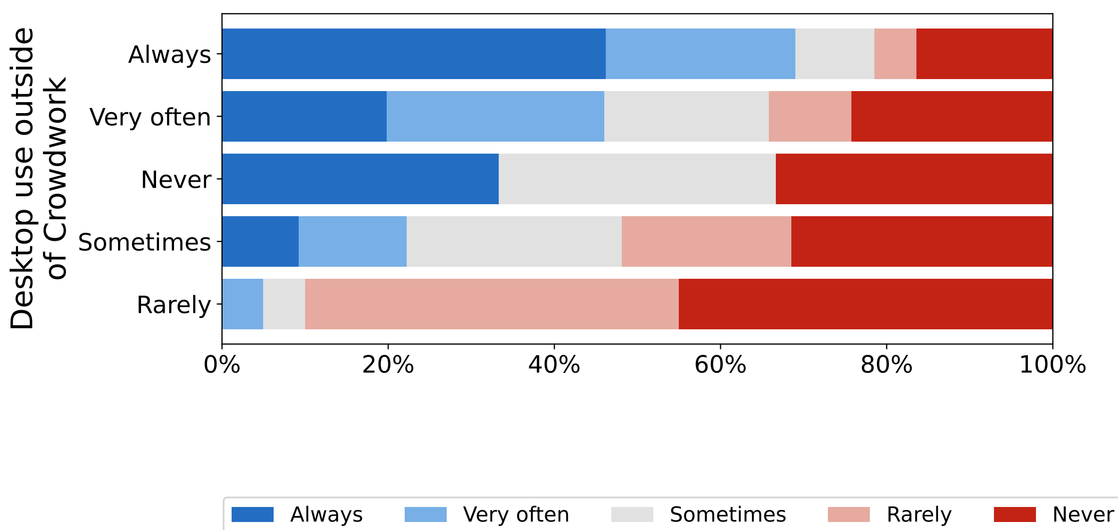


Figure 6.34: Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to currently manage crowdwork

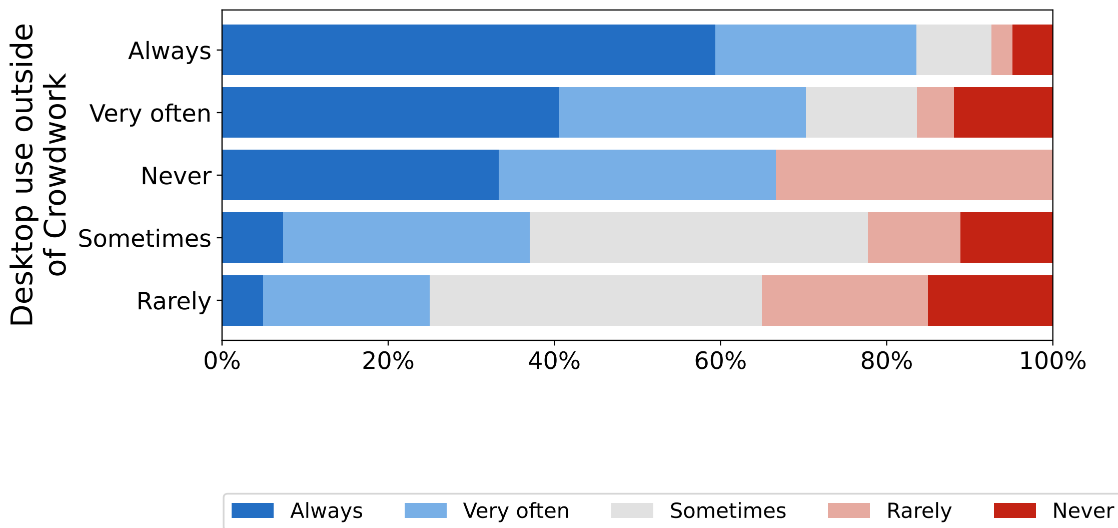


Figure 6.35: Visualization of all significant features as shown in Table 6.12 between technical factors and crowdworkers' work practices that have correlation related to desktop usage outside of crowdwork and desktop usage to manage crowdwork in ideal world

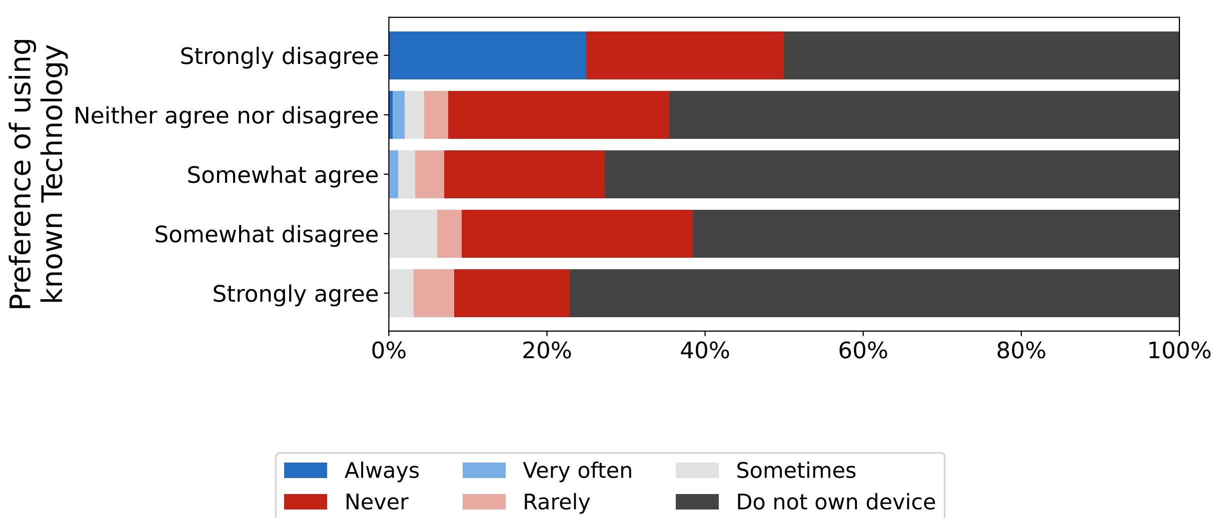


Figure 6.36: Visualization of the significant feature a shown in Table 6.12 between technical factors and crowdworkers' work practice that have correlation related to tech interest and using smart speaker to complete crowdwork in ideal world

use desktops outside of crowdwork also use them at the same frequency for completing crowdwork ($\chi^2 = 358.71$, $p < 0.001$, Corrected Effect size = 0.35). Even those who rarely use desktops outside of crowdwork still use them to some extent for completing tasks, with only a small percentage never using them (see Figure 6.32).

The current use of desktops outside of work mirrors their usage for managing crowdwork in an ideal world ($\chi^2 = 126.82$, $p < 0.001$, Corrected Effect size = 0.20) for managing and completing crowdwork ($\chi^2 = 233.70$, $p < 0.001$, Corrected Effect size = 0.28) as shown in see Figure 6.33 and Figure 6.35. Workers who frequently use desktops outside of crowdwork show a similar frequency for managing tasks in an ideal world. Interestingly, even those who rarely use desktops outside of crowdwork still prefer to use them for completing tasks in ideal world, highlighting the indispensable role of desktops in crowdwork.

While there is a strong correlation between desktop usage outside of crowdwork and for managing crowdwork, there are some exceptions. Around 15% of workers who always use desktops outside of crowdwork never use them to currently manage tasks. Conversely, the majority of workers who never use desktops outside of crowdwork do use them to manage crowdwork currently ($\chi^2 = 132.49$, $p < 0.001$, Corrected Effect size = 0.20) as shown in Figure 6.34.

Tech Interest vs Openness to Smart Speakers

A noteworthy trend is the openness to using smart speakers in an ideal world ($\chi^2 = 114.37$, $p < 0.001$, Corrected Effect size = 0.18). Workers who strongly disagree about their preference to use only familiar devices and software are more open to using smart speakers for crowdwork when high-quality tooling is available. Conversely, workers who prefer using familiar devices are less willing to adopt smart speakers, even with advanced tooling, indicating a preference for traditional devices as shown in Figure 6.36.

In summary, there is a strong correlation between the frequency of desktop usage outside of crowdwork and its use for completing and managing tasks within crowdwork. This trend persists even in an ideal scenario with high-quality tooling. Exceptions do exist, with some workers diverging from the norm, but overall, desktops remain a staple in crowdwork practices. Additionally, openness to new devices like smart speakers varies, with workers who prefer familiar devices being less likely to adopt new technology, even when advanced tools are available. These correlations highlight the importance of technical familiarity and comfort in shaping device usage among crowdworkers.

6.7.4 Correlation of Financial Factors and Crowdworkers Work Practices

We did not find any significant correlations between crowdworkers' studied financial factors such as income from crowdwork as well as the no of financial dependents as shown in Table 6.13.

6.7.5 Correlation of Profession outside of Crowdwork and Crowdworkers Work Practices

Our analysis reveals key findings on the correlation between professions outside of crowdwork and the usage of smart speaker for managing crowdwork as shown in Table 6.12 and Figure 6.37.

Profession vs Openness to Smart Speaker

Regardless of crowdworkers' professions outside of crowdwork or reliance solely on crowdwork for income, the majority of smart speaker owners do not want to use this device type for managing crowdwork, even in an ideal scenario with high-quality tooling. However, among the few who are interested in using smart speakers at least sometimes, there are notable differences based on professional background.

The highest percentage of those interested in using smart speakers come from professions in agriculture, forestry, or fishery. This group is followed by people in management, those whose only income source is from crowdwork, and individuals with professions in crafts, trades, healthcare, and IT. The chi-squared test ($\chi^2 = 81.06$, $p = 0.005$, effect size $= 0.10$) indicates a significant correlation between professional background and openness to integrating smart speakers into crowdwork practices.

Table 6.13: No significant feature that shows the correlation between financial factors and crowdworkers' work practices

Category	Feature	Statistic	Raw Effect Size	Corrected Effect Size
Financial Factors	(Financial situation of crowdworker, Use of Smart speaker to currently complete crowdwork)	$\chi^2(16, 755) = 30.1, p = 0.0175$	0.10	0.07
	(Financial situation of crowdworker,Use of Smartwatch to currently manage crowdwork)	$\chi^2(20, 755) = 31.7, p = 0.047$	0.10	0.06
	(Financial situation of crowdworker, Use of Desktop to currently manage crowdwork)	$\chi^2(20, 755) = 29.2, p = 0.084$	0.10	0.06
	(Financial situation of crowdworker, Use of Tablet to currently complete crowdwork)	$\chi^2(20, 755) = 29.0, p = 0.088$	0.10	0.05
	(Financial situation of crowdworker, Preference of Smart speaker to complete crowdwork)	$\chi^2(20, 755) = 28.8, p = 0.091$	0.10	0.05

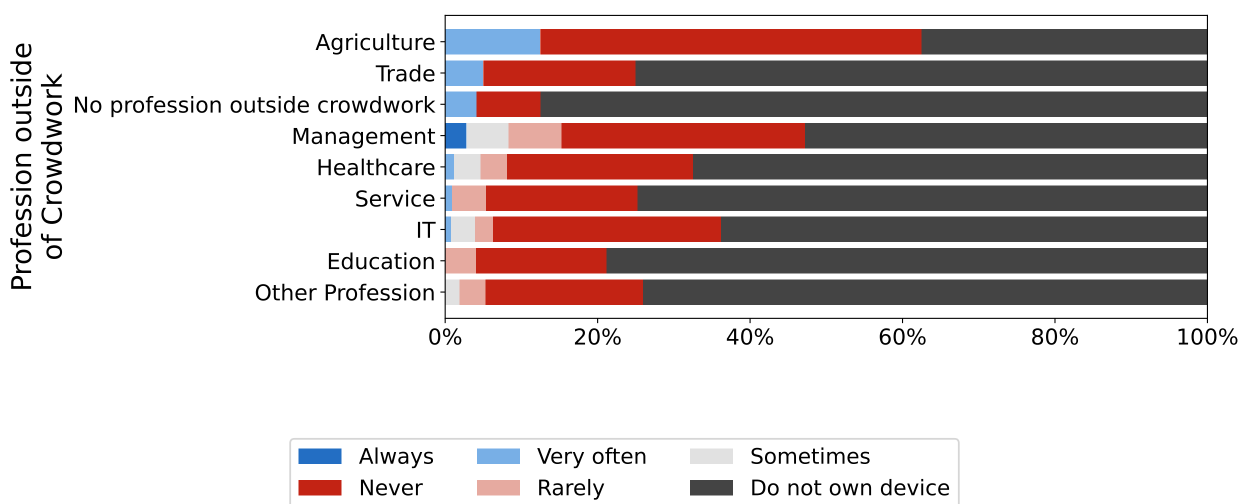


Figure 6.37: Visualization of the significant fearure as shown in Table 6.12 that shows correlation between workers' profession outside of crowdwork and smart speaker use to manage crowdwork in ideal world

In summary, while the majority of smart speaker owners, regardless of their professional backgrounds, are reluctant to use these devices for managing crowdwork, certain professions show more openness to new technologies. Specifically, individuals in agriculture, forestry, fishery, management, and those who rely solely on crowdwork, as well as those in crafts, trades, healthcare, and IT, are more willing to integrate smart speakers into their crowdwork activities. This suggests that professional background plays a role in the willingness to adopt new technologies in crowdwork practices.

6.8 Discussion

Based on our comprehensive analysis of results from Prolific and Amazon Mechanical Turk (MTurk), we propose several key implications for platform design to support the flexibility of crowdworkers. Given that MTurk and Prolific are among the most popular crowdsourcing platforms, these implications are broadly applicable to the majority of crowdsourcing environments. Additionally, specific implications reflect the unique traits of platforms that are more similar to either MTurk or Prolific.

6.8.1 Prioritize Features in Platforms for Younger Adults

Both MTurk and Prolific have a majority of users within the 25-44 age range, indicating that younger adults are the primary crowdworkers.

To prioritize features for younger adults, platforms should ensure a mobile-first design that is fully functional and optimized for mobile devices, as research indicates the significance of mobile usability for this demographic [63, 164]. For example, a mobile-first design entails creating a responsive interface that seamlessly adapts to various screen sizes, ensuring smooth navigation and interaction on smartphones and tablets. This includes features such as swipe gestures, touch-friendly buttons, and efficient load times to enhance the user experience.

Additionally, platforms should also focus on introducing nontraditional devices like smart speakers for those who prefer to try out new technologies as shown in Figure 6.36. This means integrating voice-activated features that allow users to manage tasks, receive notifications, and communicate with requesters hands-free. For instance, users can accept tasks, ask for instructions, or receive real-time updates through voice commands, making the process more accessible and convenient.

Integrating modern, tech-savvy features such as AI-powered task suggestions, gamification elements, and real-time collaboration tools can significantly enhance user engagement and motivation [36, 35]. For example, AI-powered task suggestions can help users find relevant tasks quickly, while gamification elements like achievement badges and leaderboards can make the experience more engaging. Real-time collaboration tools enable seamless communication and teamwork, enhancing productivity.

Additionally, developing easy-to-follow tutorials and interactive guides can help users, particularly older adults, become familiar with new features, addressing the onboarding challenges often faced by this group [215, 54]. These guides can include step-by-step instructions, video tutorials, and interactive walkthroughs to ensure all users can effectively utilize the platform's capabilities.

6.8.2 Tailor Task Complexity to Educational Backgrounds

Prolific attracts a more highly educated population compared to MTurk, with a higher proportion of participants holding graduate degrees.

To address these differences, platforms should tailor task complexity to educational backgrounds involves implementing a system to categorize tasks by complexity and required skills, clearly labeling tasks as beginner, intermediate, or advanced. For example, a data entry task might be labeled as beginner, a data analysis task as intermediate, and a research synthesis task as advanced. Using algorithms to match users with tasks that fit their educational background and skill levels can enhance user performance and satisfaction, as supported by studies on adaptive systems in HCI [19, 112].

Additionally, providing detailed instructions, video tutorials, and support resources for complex tasks can assist users with varying educational backgrounds. For instance, an advanced task might include a comprehensive video tutorial and a detailed step-by-step guide, ensuring that users with less experience in that area can understand and complete the task effectively. This aligns with research that emphasizes the importance of scaffolding in digital learning environments [145, 138].

By implementing these strategies, platforms can ensure that tasks are accessible and manageable for users with different educational backgrounds, enhancing overall performance and satisfaction.

6.8.3 Optimize Interfaces for Commonly Owned and Platform Specific Preferences

Both MTurk and Prolific show that workstations are the most commonly owned devices also reflected in clusters, followed by smartphone and tablet also reflected in clusters (see section 6.4). MTurk participants own more smart speakers, while Prolific users own more smartwatches, however Prolific has two groups i.e. cluster 4 who are smartwatch enthusiasts and cluster 6 who are smart speaker enthusiasts with smartphone and smartwatch reliance.

To optimize interfaces for commonly owned devices, platforms should ensure the platform is fully responsive, offering a seamless experience on desktops, smartphones, and tablets on both MTurk and Prolific. Previous research highlights the importance of responsive design in providing consistent user experiences across devices [208].

Developing device-specific features can significantly improve usability and accessibility. For instance, enhanced voice command integration for smart speakers allows users to manage tasks, receive updates, and communicate with requesters hands-free. Advanced notification systems for smartwatches enable users to get real-time alerts and updates directly on their wrist, making it easier to stay informed and responsive.

Additionally, enabling seamless synchronization of tasks and progress across all devices owned by a user is crucial for maintaining continuity and efficiency in user workflows [235]. This means that a user can start a task on their desktop, continue it on their smartphone while commuting, and finish it on their tablet at home, with all progress and changes automatically updated across all devices. This seamless cross-device interaction ensures that users can work efficiently without interruption, enhancing overall productivity and user satisfaction.

6.8.4 Support Tool Integration and Development

Supporting tool integration and development involves creating an ecosystem that supports third-party tool integration, allowing users to connect their favorite productivity tools. Research emphasizes the importance of flexible tool ecosystems in enhancing user productivity and satisfaction [76]. Our findings show that MTurk participants heavily rely on tools, as shown in Figure 4.10, with every participant using some type of tool to aid in their tasks. In contrast, a significant majority of Prolific users (85.76%) do not use any tools.

To address these differences, platforms should implement a marketplace or repository where users can discover and download tools that enhance their task performance, aligning with findings on the benefits of customizable user environments [150]. For example, creating an integrated marketplace allows users to easily access tools for task acceptance, time tracking, receiving notifications for accpeted or rejected tasks, and more, directly within the platform.

Additionally, offering training and resources to educate users about the available tools and their benefits can encourage wider adoption. This can include video tutorials, webinars, and interactive guides to help users understand how to effectively use these tools to improve their productivity. Studies have shown that user education significantly impacts technology acceptance and usage [206, 219].

By supporting a robust tool integration ecosystem, platforms can cater to the diverse needs of users on MTurk and Prolific, enhancing overall user satisfaction and productivity.

6.8.5 Tailor specific Tasks for Completion on Non-desktop Devices

There are distinct preferences in task completion across different devices. For example, tablets on MTurk are mostly used for testing software and surveys, while Prolific users favor tablets for surveys and information finding.

Tailoring task types to preferred devices includes optimizing tasks for the devices they are most commonly completed on. For instance, creating tablet-friendly surveys for both MTurk and Prolific ensures a smooth and efficient user experience. Allowing users to set device preferences and receive tasks that are best suited for those devices can enhance user experience, as supported by research on adaptive interfaces [53, 9].

For example, a user on Prolific might set their preference for using a tablet, and the platform would then prioritize sending them surveys and information-finding tasks optimized for tablet use. Continuously testing and gathering feedback on the usability of tasks across different devices ensures optimal performance and user satisfaction. This aligns with studies that emphasize the importance of iterative design and user feedback in HCI [86, 137]. By regularly updating and refining tasks based on user feedback, platforms can maintain high standards of usability and effectiveness.

6.8.6 Optimize Task Management fo Commonly mentioned Non-desktop devices

Non-desktop devices are used differently for task management on each platform. MTurk users favor active management tasks on smartphones, such as finding and accepting tasks, while Prolific users use smartphones more for notifications.

Optimizing task management for non-desktop devices involves enhancing task management features on smartphones. For instance, on MTurk, improving features such as quick access to task details, and efficient task acceptance interfaces can significantly enhance user experience. An example would be a streamlined interface that allows MTurk users to quickly scroll through available tasks, view detailed descriptions with a single tap, and accept tasks instantly.

Additionally, developing passive management tools for smartphones, such as background task tracking and automated reminders, can help workers manage their workload more effectively. For example, Prolific could implement features that allow tasks to be tracked in the background, sending automated reminders about upcoming deadlines or task status updates. This aligns with findings on the benefits of automated reminders in task management [223, 92].

By tailoring these features to the specific ways users engage with their smartphones on each platform, task management can become more efficient and user-friendly.

6.8.7 Implement Crowdwork Ecosystem Support Features

Both MTurk and Prolific platforms face ecosystem issues, such as bots stealing work from legitimate crowdworkers and not having enough time to complete crowdwork. Additionally, Prolific workers emphasize the importance of security and privacy as well as fair treatment for both requesters and workers.

Therefore, it is important to implement ecosystem support features that address these concerns. This includes developing robust anti-bot measures, such as advanced algorithms and user verification processes [111, 224], to ensure task integrity. For example, implementing CAPTCHA systems and requiring multi-factor authentication can help verify user identities and prevent bots from accessing the platform.

Enforcing policies that ensure fair treatment for both workers and requesters, such as dispute resolution mechanisms and user rating systems [190, 29], enhances trust and cooperation within the platform. For instance, introducing a transparent dispute resolution process where workers and requesters can raise concerns and have them fairly addressed can prevent misunderstandings and promote fairness.

Strengthening privacy and data protection measures by ensuring robust security protocols and user consent processes builds trust and security. Supported by research on privacy in online platforms [74, 34], this might include encrypting user data and providing clear information about data usage, as well as obtaining explicit user consent for data collection and processing.

By implementing these ecosystem support features, both MTurk and Prolific can improve the integrity, trust, and overall user experience on their platforms.

6.8.8 Region Specific Design for Platforms

Across all regions, Africa and Canada show the highest desktop usage for managing and completing crowdwork tasks, even with high-quality tooling. Oceania leads in smartphone usage for managing tasks, while Europe has the lowest smartphone usage for task completion. These correlations highlight the diverse device preferences and work practices across different demographics and regions among crowdworkers. The findings suggest significant regional differences in device usage for crowdwork, with Canada and Africa favoring desktops, and Oceania, UK Africa and USA showing a strong preference for smartphones.

Platforms should prioritize robust desktop functionalities for Africa and Canada, ensuring stability, efficiency, and offline capabilities [130]. For example, in Africa, where internet connectivity can be unstable, providing offline capabilities allows workers to continue their tasks without interruption. In Canada, optimizing software for efficiency and stability helps crowdworkers handle large datasets and complex tasks seamlessly.

For Oceania, UK, Africa, and USA, it is important to focus on optimizing mobile interfaces with intuitive, responsive applications [210, 102]. Especially in the USA and UK, where the number of crowdworkers is the highest, platforms should offer features like real-time notifications, voice-activated commands, and streamlined task management to enhance the mobile user experience. For example, a responsive mobile interface that adapts to different screen sizes and offers easy navigation can significantly improve task efficiency and user satisfaction.

Overall the findings underscore the importance of adopting a user-centered approach in the design and development of crowdwork platforms. By understanding and addressing the diverse technological needs and preferences of crowdworkers across different regions, backgrounds, device prefereces platforms will not only enhance to support the flexibility of crowdworkers but also enable platforms to remain competitive in the evolving gig economy landscape.

6.9 Limitation

This study has several limitations that should be considered when interpreting the findings. Firstly, the findings cannot be generalized to all crowdsourcing platforms beyond those represented in the survey sample. Our study primarily focuses on platforms like Amazon Mechanical Turk and Prolific, and the results may not be applicable to other platforms with different structures and user bases. The study does not include specific performance metrics, such as task completion time or accuracy, for tasks completed and managed using different devices. This omission means we cannot assess the impact of device type on individual performance outcomes. The study does not provide insights into how device usage practices, task traits, or ecosystem traits might change over time. Future research could benefit from a longitudinal approach to understand the evolution of these factors.

Chapter 7

Incentives of Platform Owners , Potential Risk and Limitation

7.1 Financial Incentives of Platform Owners vs. Traits to Support Worker Flexibility

Many crowdsourcing platform companies are now collecting data either for their own use or on behalf of clients developing large language models (LLMs). This data collection is aimed at training and evaluating LLM responses, with a focus on obtaining diverse perspectives rather than relying on majority voting [46]. Consequently, platform owners are incentivized to gather data from different regions to ensure a wide range of viewpoints. However, people in different regions often do not own a variety of devices as shown in Chapter 6, which limits the diversity of data collected. Platform owners are therefore interested in including diverse devices to attract more participants, enhancing data diversity and drawing more requesters to their platforms. This increases financial revenue for the platform owners.

However, this financial incentive may lead platform owners to limit some features that support worker flexibility, such as blocking bad requesters who never respond to workers or providing impactful work. The focus on maximizing user numbers and financial gains might overshadow the implementation of measures that could better support the flexibility and well-being of crowdworkers.

7.2 Potential Risk for Using Various Devices in Crowdwork

While using various devices for crowdwork can enhance flexibility by allowing workers to choose their preferred devices based on their situations, this convenience comes with potential risks. Imagine the allure of being able to manage and complete tasks from anywhere, anytime—whether at a café, a park, or even while traveling. However, this very flexibility can blur the boundaries between work and personal life, leading to overwork and stress. Picture a crowdworker trying to catch up on tasks during a long commute, glancing at their device while driving. This dangerous multitasking can lead to serious accidents, turning the promise of flexibility into a hazard. Therefore, it is crucial for crowdworkers to use these devices responsibly, maintaining a healthy separation between their work and personal lives. Crowdsourcing platforms, only allow workers who are over 18 years old, therefore we expect adult crowdworkers to exercise responsible usage of devices to mitigate these risks.

7.3 Limitation

This thesis has several limitations. Firstly, it focuses on only four non-workstation devices: smartphone, tablet, smart speaker, and smartwatch. Future studies should expand the scope to include a broader range of devices, such as VR headsets to gain a more comprehensive understanding of crowdworkers' device practices and preferences to support the flexible crowdwork enviornment.

Secondly, this thesis examines crowdworkers' current device practices and preferences exclusively on Amazon Mechanical Turk (MTurk) and Prolific. Researchers should explore a wider variety of crowdsourcing platforms beyond MTurk and Prolific, such as Upwork, CrowdFlower, and other gig economy platforms. This would help to identify platform-specific differences and provide a more holistic view of the crowdsourcing landscape.

Finally, this thesis is limited to a specific timeframe to understand crowdworkers' current practices and device preferences theoretically. Future studies should consider longitudinal research to track changes in device usage and preferences over time, as well as develop prototypes or systems to understand the impact of devices on crowdwork practices in real-world scenarios.

Chapter 8

Discussion

8.1 Broader Implications in the Workflow of Society

This research provides a comprehensive understanding of crowdworkers' task completion and management practices across various non-workstation devices, such as smartphones, tablets, speakers, and smartwatches. While the primary focus is on enhancing the flexibility and support for crowdworkers, the findings have broader implications that extend into the wider societal workflow, particularly in the context of digital device usage and remote work dynamics.

8.1.1 Impact on Flexibility on Different Type of Profession

Prior research indicates that professionals can benefit from increased fleixbility [105]. Exploring the versatility of flexibility reveals its unique impact across diverse professions, underscoring that one size does not fit all when it comes to implementing flexible work arrangements [172, 7]. Our work exploring the versatility of flexibility within crowdwork reveals its unique impact across diverse profession. This perspective acknowledges that while flexibility is universally beneficial, its optimal application requires a tailored approach that considers the specific needs and challenges of each profession as discussed in Chapter 5. Whether it's healthcare's variable shift patterns, hybrid teaching model in education, the tech industry's remote work opportunities, or retail's shift swapping, each method underscores how tailored flexibility improves work-life balance. This approach respects the distinct requirements of various job roles and places a premium on employee well-being and satisfaction, illustrating that well-considered flexibility is key to enhancing professional life across sectors.

8.1.2 Informing Flexible and Supportive Work Policies

As workforce fleixbility becomes more prevalent, there is a critical need to update labor laws and social protections to align with the realities of contemporary work arrangements. Ensuring that all workers, irrespective of their employment status, have equitable access to benefits is fundamental to fostering a fair and inclusive work environment. This also necessiates legislative adjustments to guarantee that all workers including the gig workers receive the same rights and protection as those stationed in traditional office environments such as child and family benefits, maternity protection, unemployment support, employment injury benefits, sickness benefits, health protection and many more [213, 99].

The insights gained from understanding crowdworkers' practices and expectations can inform remote work policies at an organizational level. By recognizing the diverse needs of their employees , organizations can implement policies that promote work-life balance, such as flexible working hours and the option to use multiple devices for task completion. For instance, companies can invest in technology that supports cross-device workflows, ensuring that employees have access to necessary tools regardless of their location. Additionally, by understanding the preferences for different device usage, organizations can provide appropriate training and resources to help employees optimize their productivity. These practices can lead to increased job satisfaction, productivity, and retention, as employees feel more empowered to manage their work in ways that best suit their individual preferences and lifestyles.

8.1.3 Economic Development and Innovation

Entrepreneurship is crucial in defining the economic landscape of nations. It acts as the lifeblood of the market, infusing employment and growth into the economy. Entrepreneurs, with their bold visions, not only open up new avenues of possibility that promote capital formation and generate widespread employment opportunities but also drive innovation and support new business ideas [38, 199]. This creates a cascading effect that not only transforms individual lives but also bolsters the overall economy. In today's rapidly changing market, there is a pressing need to support a dynamic economy [168]. By fostering entrepreneurial initiatives, we can ensure that the economy remains resilient, adaptable, and capable of sustaining long-term growth [48].

Our work supporting the flexibility of crowdworkers highlights a more dynamic economy. By allowing individuals to engage with various projects or transition between roles fluidly, it increases the labor market's ability to quickly respond to economic changes. This agility facilitates faster cycles of innovation, as companies can access a global talent pool on demand, minimizing the delay between ideation and execution. Additionally, this flexibility encourages entrepreneurship by providing individuals the opportunity to pursue new business ventures with lower risk and less overhead.

Overall, while this thesis focuses on crowdworkers, the implications of the findings extend far beyond, offering valuable lessons for fleixbility on different type of profession , informing flexible and supportive work policies and offers significant benefits for economic development and innovation. These contributions can significantly impact various sectors, ultimately fostering a more flexible, efficient, and inclusive work environment in society.

8.2 Implications in the Future of AI supported Crowdworld

The findings of this thesis have several significant implications for the future of crowdwork particularly in enhancing the flexibility and efficiency of crowdworkers through the integration of various device types, with a strong emphasis on the role of Artificial Intelligence (AI).

8.2.1 AI-Driven Workflow Optimization

The research highlights the potential for AI to tailor workflows to specific device types. AI can analyze the unique affordances and constraints of different devices, such as smartphones, tablets, smart speakers, and smartwatches, to optimize task completion and management systems. By leveraging AI algorithms to understand device-specific user behaviors and preferences, platforms can offer adaptive interfaces and functionalities that enhance productivity and user satisfaction [173, 198].

8.2.2 AI-Driven Scripts and Tool Unification for Efficiency

The desire across groups of workers for a unified scripting approach where a master script replaces multiple device-specific scripts presents an opportunity for AI to streamline task management processes [167]. By employing AI to automate and optimize task assignment, acceptance, and completion processes, crowdworkers can engage more efficiently. AI can also develop adaptive filtering algorithms for HITs and requesters, enhancing task relevance and quality, echoing the needs of crowdworkers as showin in the thesis.

8.2.3 Cross-Device AI Integration for Enhanced Task Management

The transition observed in different groups of crowdworkers towards using a broader range of devices, from smartphones and tablets to wearables like smart speakers and watches, calls for AI systems capable of seamless cross-device integration. AI could provide sophisticated solutions allowing for smooth work transition across devices without loss of context or productivity. This involves developing AI algorithms to synchronize work progress, manage notifications effectively, and ensure tasks are suited to each device's capabilities [85, 179].

8.2.4 Personalized AI Support Systems

By identifying systematic differences in crowdworkers' practices and expectations across various devices, the research paves the way for creating AI-driven, customizable support systems. AI can provide personalized assistance by learning from individual worker's interactions and preferences, offering context-aware recommendations, task prioritization, and real-time support tailored to the worker's preferred device. This personalization can reduce cognitive load and improve task efficiency, leading to better performance and higher job satisfaction.

8.2.5 AI-Powered Platform Insights

The exploration of platform biases, particularly comparing Amazon Mechanical Turk (MTurk) and Prolific, highlights how AI can be utilized to analyze and understand platform-specific traits that influence crowdwork practices. For instance, AI-driven analytics can examine task design constraints, user demographics, and interaction patterns on each platform [197] . By analyzing these factors, AI can identify differences in how tasks are completed, the types of tasks preferred by workers, and the overall user experience. This deep analysis allows platform designers to create more supportive and adaptive environments that cater to diverse crowdworker needs. Understanding these platform-specific nuances can inform the development of best practices and features, ultimately enhancing the user experience and ensuring fairness and efficiency across different crowdsourcing platforms.

8.2.6 AI driven Solutions to Prevent Bots in Crowdwork

Bots on crowdsourcing platforms undermine the integrity of the work and unfairly compete with legitimate crowdworkers, as shown in the findings of the thesis. Advanced AI solutions can mitigate this issue by identifying and preventing bot activity, ensuring a fairer environment. AI can detect bots by analyzing repetitive and predictable patterns in user activity.

However, it is important to acknowledge the distinction between old bots and new bots. Old bots exhibited predictable and repetitive patterns, making them relatively easier to detect with traditional AI algorithms. These bots often followed simple, rule-based behaviors that could be identified through straightforward pattern recognition techniques.

Generative AI fundamentally changes the nature of bot detection solutions. Generative AI can create more sophisticated bots that mimic human behavior with high fidelity, producing nuanced and less predictable patterns that closely resemble genuine human users. This increased sophistication makes detecting bots much harder for traditional AI systems, as these advanced bots can adapt and evolve their behaviors to avoid detection.

To address the challenges introduced by generative AI, more advanced and adaptive solutions are required. AI-driven CAPTCHAs that adapt in complexity can effectively differentiate between bots and humans [133]. AI systems must continuously update to detect new bot behaviors, incorporating the latest data and feedback from human moderators. AI-powered multi-factor authentication (MFA) can detect anomalies in login patterns and device usage, prompting additional verification for suspicious activity. Continuously updating AI models with new data on bot behaviors and integrating advanced techniques like deep learning can handle sophisticated bot behavior [209, 79].

Using feedback from human moderators to refine detection algorithms ensures the effectiveness of AI solutions. Additionally, ensuring that AI solutions minimize false positives and do not unfairly flag legitimate crowdworkers is crucial for maintaining transparency and trust. While generative AI introduces new challenges in bot detection, a combination of adaptive techniques, continuous updates, and human oversight can help maintain the integrity of crowdsourcing platforms.

8.2.7 Enhanced Communication and Design through AI

The interest in better communication tools and task design across devices, particularly suggests a need for AI that supports more intuitive and effective interactions between crowdworkers and requesters. AI could facilitate real-time, context-aware communication, ensuring messages and instructions are clearly conveyed and understood, regardless of the device used and automated routine processes such as finding, queuing, and accepting tasks, thereby freeing workers from the drudgery of manual task management [237] . In addition to that, AI can assist in optimizing task interfaces for smaller screens or voice-controlled devices, making crowdwork more accessible to individuals who rely on non-workstation devices. This inclusivity can open up new opportunities for a wider range of crowdworkers, contributing to a more equitable and diverse workforce [193].

8.2.8 AI Supported Equal Work Opportunity for All Crowdworkers

In rural areas, crowdworkers face challenges like limited internet connectivity.This can hinder their ability to access and complete tasks, leading to difficulties in onboarding and impacting their earning opportunities [30, 55, 216]. These challenges highlight the need for inclusive

design to ensure equitable access and opportunities for all workers, regardless of their location
.

To address these issues, several key solutions are essential. Developing offline-capable AI tools that sync when online and optimizing applications for low-bandwidth environments can reduce data usage. Creating lightweight, mobile-friendly interfaces and implementing edge computing for local processing can further enhance accessibility. Supporting multiple languages and localized content improves usability. Establishing community hubs with reliable internet and shared resources can provide crucial support. Using AI to allocate tasks based on connectivity ensures that low-data tasks are assigned to those with limited access. Offering training and support helps rural workers use AI tools effectively. These solutions will make AI-supported work more inclusive, ensuring equal opportunities for all crowdworkers.

8.2.9 Ethical and Privacy Considerations

While the integration of AI into the crowdwork ecosystem offers exciting possibilities, it is imperative to navigate these advancements with caution. The push towards greater automation and personalization brings to the fore ethical and privacy concerns that must be addressed. Ensuring the ethical deployment of these technologies involves a commitment to transparency, data privacy, and the safeguarding of workers' rights, ensuring that the future of crowdwork remains not only innovative but also respectful and secure.

The thesis identifies several avenues for future research, including exploring advanced AI-driven solutions to further optimize task management and completion processes across various devices. Future studies can investigate how AI can enhance the adaptability of task interfaces, improve real-time feedback mechanisms, equal work opportunity and support dynamic workflow adjustments based on real-time data. This can lead to the development of more intelligent and responsive crowdwork platforms.

Chapter 9

Conclusion

In this thesis, we have provided a new understanding of traits related to the stages of crowdwork, including task management and task completion using four non-workstation device types: smartphone, tablet, smart speaker, and smartwatch. Additionally, we have examined crowdwork ecosystem traits that support the overall flexibility of crowdworkers.

Our results, presented in Chapter 3, suggest that to support the growing interest of crowdworkers in mobile devices, designers should follow our proposed taxonomy of characteristics.

By understanding the current and desired practices of crowdworkers, our findings offer a set of recommendations for crowdwork platforms, tools, and requesters (i.e., researchers) to better support task completion and management on non-workstation devices. These recommendations, detailed in Chapter 4, aim to provide crowdworkers with more choices regarding where and how they complete their work, thereby giving them greater flexibility.

Furthermore, by examining the systematic differences in crowdworkers' work practices based on the stage of work and device type in Chapter 5, our findings offer insights for platform designers to create more personalized, flexible, and efficient crowdsourcing platforms for crowdworkers.

Finally, to validate the generalizability of our findings from MTurk, we conducted another large-scale quantitative survey on Prolific. Our findings provide guidance to platform designers to support the flexibility of crowdworkers on two majorly used platforms a shown in Chapter 6.

In conclusion, this thesis contributes to a deeper understanding of the traits (see Figure 9.1)influencing crowdwork on non-workstation devices and offers practical recommendations for enhancing the flexibility and efficiency of crowdworking platforms. By addressing these traits, we can create a more supportive environment for crowdworkers, ultimately benefiting both the workers and the platforms they use.

Chapters	Focus	Worker traits	Task traits	Ecosystem traits	Clustering workers	Device challenges	Devices studied	Stage of work
Chapter 3	Task usability	☑	☑			☑	Smartphone	Completion
Chapter 4	Work practices	☑	☑	☑		☑	Workstation, Tablet, Smartphone, Smart speaker, Watch	Completion, Management
Chapter 5	Systematic difference between workers		☑	☑	☑	☑	Workstation, Tablet, Smartphone, Smart speaker, Watch	Completion, Management
Chapter 6	Prolific workers, comparison to MTurk	☑	☑	☑	☑	☑	Workstation, Tablet, Smartphone, Smart speaker, Watch	Completion, Management

Figure 9.1: Overview of studied traits

Bibliography

[1] Abououf, M., Otrok, H., Mizouni, R., Singh, S., and Damiani, E. (2020). How artificial intelligence and mobile crowd sourcing are inextricably intertwined. *IEEE Network*, 35(3):252–258. 1

[2] Acer, U. G., Broeck, M. v. d., Forlivesi, C., Heller, F., and Kawsar, F. (2019). Scaling crowdsourcing with mobile workforce: A case study with belgian postal service. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2):1–32. 3, 12, 51, 78

[3] Agarwal, R. and Venkatesh, V. (2002). Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability. *Information systems research*, 13(2):168–186. 46

[4] Aggarwal, D., Khot, R. A., and Yi, J.-Y. (2024). Zenscape: Encouraging microbreaks from screen-based work through a tangible system. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–17. 97

[5] Ahire, S., Rohs, M., and Benjamin, S. (2022). Ubiquitous work assistant: Synchronizing a stationary and a wearable conversational agent to assist knowledge work. In *2022 Symposium on Human-Computer Interaction for Work*, pages 1–9. 97

[6] Ahmad, N., Rextin, A., and Kulsoom, U. E. (2018). Perspectives on usability guidelines for smartphone applications: An empirical investigation and systematic literature review. *Information and Software Technology*, 94:130–149. 72

[7] Ahmadi, S., Jansen, J. J., and Eggers, J. (2022). Using stretch goals for idea generation among employees: One size does not fit all! *Organization Science*, 33(2):671–687. 175

[8] Aipe, A. and Gadiraju, U. (2018). Similarhits: Revealing the role of task similarity in microtask crowdsourcing. In *Proceedings of the 29th on Hypertext and Social Media*, pages 115–122. 21, 22

[9] Akiki, P. A., Bandara, A. K., and Yu, Y. (2014). Adaptive model-driven user interface development systems. *ACM Computing Surveys (CSUR)*, 47(1):1–33. 167

[10] Alabduljabbar, R. and Al-Dossari, H. (2019). A dynamic selection approach for quality control mechanisms in crowdsourcing. *IEEE Access*, 7:38644–38656. 25, 99

[11] Anthony, L., Brown, Q., Nias, J., Tate, B., and Mohan, S. (2012). Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 225–234. 28

[12] Baltes, B. B., Briggs, T. E., Huff, J. W., Wright, J. A., and Neuman, G. A. (1999). Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *Journal of applied psychology*, 84(4):496. 2, 12

[13] Bernaerts, Y., Druwé, M., Steensels, S., Vermeulen, J., and Schöning, J. (2014). The office smartwatch: development and design of a smartwatch app to digitally augment interactions in an office environment. In *Proceedings of the 2014 companion publication on Designing interactive systems*, pages 41–44. 11

[14] Bernstein, M. S., Karger, D. R., Miller, R. C., and Brandt, J. (2012a). Analytic methods for optimizing realtime crowdsourcing. *arXiv preprint arXiv:1204.2995*. 21

[15] Bernstein, M. S., Teevan, J., Dumais, S., Liebling, D., and Horvitz, E. (2012b). Direct answers for search queries in the long tail. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 237–246. 21, 22

[16] Blum, J. R., Frissen, I., and Cooperstock, J. R. (2015). Improving haptic feedback on wearable devices through accelerometer measurements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 31–36. 73

[17] Braun, V. and Clarke, V. (2012). Thematic analysis. 25

[18] Brewer, R., Pierce, C., Upadhyay, P., and Park, L. (2022). An empirical study of older adult's voice assistant use for health information seeking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 12(2):1–32. 10

[19] Brusilovsky, P. and Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web: methods and strategies of web personalization*, pages 3–53. Springer. 165

[20] Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5. 1

[21] Calacci, D. (2022). Organizing in the end of employment: information sharing, data stewardship, and digital workerism. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 12

[22] Carter-Browne, B. M., Paletz, S., Campbell, S. G., Carraway, M. J., Vahlkamp, S. H., Schwartz, J., and O'Rourke, P. (2021). There is no "ai" in teams: A multidisciplinary framework for ais to work in human teams. *Applied Research Laboratory for Intelligence and Security (ARLIS) Report June*. 96

[23] Cecchinato, M. E., Cox, A. L., and Bird, J. (2015). Working 9-5? professional differences in email and boundary management practices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3989–3998. 96

[24] Center, P. R. (2021). Mobile fact sheet. *Pew Research Center*. 9, 49

[25] Chandra, A., Chandna, P., Deswal, S., and Kumar, R. (2009). Ergonomics in the office environment: a review. In *Proceedings of the international Conference of Energy and Environment. Chandigarh, Haryana, India*. 7

[26] Chi, P.-Y., Batra, A., and Hsu, M. (2018). Mobile crowdsourcing in the wild: Challenges from a global community. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct*, pages 410–415. 11, 14, 51, 78

[27] Chopra, M., Medhi Thies, I., Pal, J., Scott, C., Thies, W., and Seshadri, V. (2019). Exploring crowdsourced work in low-resource settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13. 11

[28] Churchill, E. F. (2018). Putting accessibility first. *ACM/SIGCSE Seek*, 25:24. 47

[29] Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., Vincent, N., and Li, H. (2022). Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 778–788. 169

[30] Corps, M. (2019). Towards a digital workforce: understanding the building blocks of kenya's gig economy. final report. *Nairobi, Kenya*. 180

[31] de Lima Salgado, A., do Amaral, L. A., de Mattos Fortes, R. P., Chagas, M. H. N., and Joyce, G. (2017). Addressing mobile usability and elderly users: Validating contextualized heuristics. In *International Conference of Design, User Experience, and Usability*, pages 379–394. Springer. 15

[32] Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Scott, Foresman,. 17

[33] Della Mea, V., Maddalena, E., and Mizzaro, S. (2013). Crowdsourcing to mobile users: A study of the role of platforms and tasks. In *DBCrowd*, pages 14–19. Citeseer. 25, 99

[34] Demertzi, V., Demertzis, S., and Demertzis, K. (2023). An overview of privacy dimensions on the industrial internet of things (iiot). *Algorithms*, 16(8):378. 169

[35] Deterding, S. (2019). Gamification in management: Between choice architecture and humanistic design. *Journal of Management Inquiry*, 28(2):131–136. 165

[36] Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. 165

[37] Dey, A. K., Abowd, G. D., and Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human–Computer Interaction*, 16(2-4):97–166. 74

[38] Dhaliwal, A. (2016). Role of entrepreneurship in economic development. *International Journal of scientific research and management*, 4(6):4262–4269. 177

[39] Difallah, D., Filatova, E., and Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18*, pages 135–143, New York, New York, USA. ACM Press. 1

[40] Duggan, E. W. (2003). Generating systems requirements with facilitated group techniques. *Human-Computer Interaction*, 18(4):373–394. 17

[41] Dündar, H. and Akçayır, M. (2014). Implementing tablet pcs in schools: Students' attitudes and opinions. *Computers in Human Behavior*, 32:40–46. 10

[42] Dunn, M., Munoz, I., and Jarrahi, M. H. (2023). Dynamics of flexible work and digital platforms: Task and spatial flexibility in the platform economy. *Digital Business*, 3(1):100052. 2

[43] Durward, D., Blohm, I., and Leimeister, J. M. (2016). Is there papa in crowd work?: A literature review on ethical dimensions in crowdsourcing. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, pages 823–832. IEEE. 104

[44] Dutta, S., Linder, R., Lowe, D., Rosenbalm, R., Kuzminykh, A., and Williams, A. C. (2022a). Mobilizing crowdwork: A systematic assessment of the mobile usability of hits. In

CHI Conference on Human Factors in Computing Systems, pages 1–20. 3, 8, 51, 52, 71, 77, 78

[45] Dutta, S., Linder, R., Ruoti, S., Williams, A., and Kuzminykh, A. (2022b). Beyond a one-size-fits-all approach: Towards personalizing multi-device setups in crowdwork. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pages 30–31. 51, 77

[46] Dutta, S., Mittal, S., Chen, S., Ramachandran, D., Rajakumar, R., Kivlichan, I., Mak, S., Butryna, A., and Paritosh, P. (2023). Modeling subjectivity (by mimicking annotator annotation) in toxic comment identification across diverse communities. *arXiv preprint arXiv:2311.00203*. 172

[47] Eagle, N. (2009). txteagle: Mobile crowdsourcing. In *International Conference on Internationalization, Design and Global Development*, pages 447–456. Springer. 11

[48] Economic Innovation Group (2024). The case for economic dynamism. Accessed: 2024-05-30. 177

[49] Ekandjo, T. (2024). *Human-AI Interaction In Regulating Productivity And Wellbeing*. PhD thesis, Open Access Te Herenga Waka-Victoria University of Wellington. 96

[50] Faste, H., Rachmel, N., Essary, R., and Sheehan, E. (2013). Brainstorm, chainstorm, cheatstorm, tweetstorm: new ideation strategies for distributed hci design. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 1343–1352. 17

[51] Feng, Y., Cao, Q., Qi, H., and Ruoti, S. (2020). Sencaptcha: A mobile-first captcha using orientation sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–26. 74

[52] Ferguson, J. M. (2017). Middle school students' reactions to a 1: 1 ipad initiative and a paperless curriculum. *Education and Information Technologies*, 22(3):1149–1162. 10

[53] Findlater, L. and Gajos, K. Z. (2009). Design space and evaluation challenges of adaptive graphical user interfaces. *AI Magazine*, 30(4):68–68. 167

[54] Fisk, A. D., Czaja, S. J., Rogers, W. A., Charness, N., and Sharit, J. (2020). *Designing for older adults: Principles and creative human factors approaches*. CRC press. 165

[55] Flores-Saviaga, C., Li, Y., Hanrahan, B., Bigham, J., and Savage, S. (2020). The challenges of crowd workers in rural and urban america. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 8, pages 159–162. 2, 47, 49, 180

[56] Fuchs, A., Sturdee, M., and Schöning, J. (2018). Foldwatch: using origami-inspired paper prototypes to explore the extension of output space in smartwatches. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, pages 47–59. 8

[57] Gadiraju, U., Checco, A., Gupta, N., and Demartini, G. (2017). Modus operandi of crowd workers: The invisible role of microtask work environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–29. 8

[58] Gadiraju, U., Kawase, R., and Dietze, S. (2014). A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 218–223. 3, 25, 77, 99

[59] Gajendran, R. S. and Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: meta-analysis of psychological mediators and individual consequences. *Journal of applied psychology*, 92(6):1524. 2

[60] Garfield, M. J. (2005). Acceptance of ubiquitous computing. *Information Systems Management*, 22(4):24–31. 10

[61] Garg, R. and Moreno, C. (2019). Exploring everyday sharing practices of smart speakers. In *IUI Workshops*. 75

[62] Garg, R. and Sengupta, S. (2020). He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–24. 10

[63] Ghose, A. and Han, S. P. (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science*, 57(9):1671–1691. 164

[64] Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems*, 12(4):436–445. 55, 104

[65] Glodek, M., Honold, F., Geier, T., Krell, G., Nothdurft, F., Reuter, S., Schüssel, F., Hörnle, T., Dietmayer, K., Minker, W., et al. (2015). Fusion paradigms in cognitive technical systems for human–computer interaction. *Neurocomputing*, 161:17–37. 94

[66] Goncalves, J., Feldman, M., Hu, S., Kostakos, V., and Bernstein, A. (2017a). Task routing and assignment in crowdsourcing based on cognitive abilities. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1023–1031. 3, 51, 78

[67] Goncalves, J., Ferreira, D., Hosio, S., Liu, Y., Rogstadius, J., Kukka, H., and Kostakos, V. (2013). Crowdsourcing on the spot: altruistic use of public displays, feasibility, performance, and behaviours. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 753–762. 11

[68] Goncalves, J., Hosio, S., Rogstadius, J., Karapanos, E., and Kostakos, V. (2015). Motivating participation and improving quality of contribution in ubiquitous crowdsourcing. *Computer networks*, 90:34–48. 51, 78

[69] Goncalves, J., Hosio, S., Van Berkel, N., Ahmed, F., and Kostakos, V. (2017b). Crowdpickup: Crowdsourcing task pickup in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–22. 74

[70] Goncalves, J., Kukka, H., Sánchez, I., and Kostakos, V. (2016). Crowdsourcing queue estimations in situ. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1040–1051. 11

[71] Good, B. M. and Su, A. I. (2013). Crowdsourcing for bioinformatics. *Bioinformatics*, 29(16):1925–1933. 21

[72] Goth, G. (1999). Mobile devices present integration challenges. *IT professional*, 1(3):11–15. 96

[73] Grady, C. and Lease, M. (2010). Crowdsourcing document relevance assessment with mechanical turk. In *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's mechanical turk*, pages 172–179. 21, 22

[74] Grobler, M., Gaire, R., and Nepal, S. (2021). User, usage and usability: Redefining human centric cyber security. *Frontiers in big Data*, 4:583723. 169

[75] Grudin, J. and Poltrock, S. (2012). Taxonomy and theory in computer supported cooperative work. *The Oxford handbook of organizational psychology*, 2:1323–1348. 46

[76] Guimbretière, F., Stone, M., and Winograd, T. (2001). Fluid interaction with high-resolution wall-size displays. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 21–30. 167

[77] Guo, X., Liu, J., and Chen, Y. (2017). Fitcoach: Virtual fitness coach empowered by wearable mobile devices. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE. 10

[78] Gupta, A., Thies, W., Cutrell, E., and Balakrishnan, R. (2012). mclerk: enabling mobile crowdsourcing in developing regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1843–1852. 11

[79] Gupta, D. (2023). From phishing to fraud: How ai can safeguard your customers. 180

[80] Haliburton, L., Wozniak, P. W., Schmidt, A., and Niess, J. (2021). Charting the path: Requirements and constraints for technology-supported walking meetings. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–31. 7

[81] Hanrahan, B. V., Chen, A., Ma, J., Ma, N. F., Squicciarini, A., and Savage, S. (2021). The expertise involved in deciding which hits are worth doing on amazon mechanical turk. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23. 47, 49

[82] Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–14. 25

[83] Hassler, E., Carver, J. C., Kraft, N. A., and Hale, D. (2014). Outcomes of a community workshop to identify and rank barriers to the systematic literature review process. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10. 17, 19

[84] Hegedus, D. and Rasmussen, R. (1986). Task effectiveness and interaction process of a modified nominal group technique in solving an evaluation problem. *Journal of Management*, 12(4):545–560. 18

[85] Heinimaki, T. J. and Vanhatupa, J.-M. (2013). Implementing artificial intelligence: a generic approach with software support. *Proceedings of the Estonian Academy of Sciences*, 62(1). 178

[86] Henderson, A. (2002). Interaction design: beyond human-computer interaction. *Ubiquity*, 2002(March):2. 168

[87] Hettiachchi, D., Sarsenbayeva, Z., Allison, F., van Berkel, N., Dingler, T., Marini, G., Kostakos, V., and Goncalves, J. (2020a). " hi! i am the crowd tasker" crowdsourcing through digital voice assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14. 3, 12, 48, 51, 78

[88] Hettiachchi, D., van Berkel, N., Dingler, T., Allison, F., Kostakos, V., and Goncalves, J. (2019). Enabling creative crowd work through smart speakers. In *Workshop on Designing Crowd-powered Creativity Support Systems (CHI'19 Workshop)*, pages 1–5. 8, 12, 51, 78

[89] Hettiachchi, D., Wijenayake, S., Hosio, S., Kostakos, V., and Goncalves, J. (2020b). How context influences cross-device task acceptance in crowd work. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 8, pages 53–62. 2, 6, 7, 12, 14, 21, 22, 25, 48, 51, 52, 53, 74, 78, 99

[90] Hosio, S., Goncalves, J., Lehdonvirta, V., Ferreira, D., and Kostakos, V. (2014). Situated crowdsourcing using a market model. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 55–64. 11, 51, 78

[91] Hosseini, M., Phalp, K., Taylor, J., and Ali, R. (2014). The four pillars of crowdsourcing: A reference model. In *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*, pages 1–12. IEEE. 21

[92] HR, M. H. (2015). Automation. 168

[93] Huang, Y.-M., Shadiev, R., Sun, A., Hwang, W.-Y., and Liu, T.-Y. (2017). A study of the cognitive diffusion model: facilitating students' high level cognitive processes with authentic support. *Educational Technology Research and Development*, 65(3):505–531. 10

[94] Husmann, M., Murolo, A., Kick, N., Di Geronimo, L., and Norrie, M. C. (2018). Supporting out of office software development using personal devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11. 48

[95] Hutchings, H. M. and Pierce, J. S. (2006). Understanding the whethers, hows, and whys of divisible interfaces. In *Proceedings of the working conference on Advanced visual interfaces*, pages 274–277. 48

[96] Huws, U. and Joyce, S. (2016). Crowd working survey: size of the UK's 'Gig Economy' revealed for the first time. Technical report. 1

[97] Huws, U. and Joyce, S. (2017). First survey results reveal high levels of crowd work in Switzerland. Technical report. 1

[98] Ikeda, K. and Hoashi, K. (2017). Crowdsourcing go: Effect of worker situation on mobile crowdsourcing performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1142–1153. 21, 22

[99] International Labour Organization (2023). Title of the ilo report. `https://www.ilo.org/wcmsp5/groups/public/---dgreports/---ddg_p/documents/publication/wcms_867535.pdf`. 176

[100] Iqbal, S. T., Teevan, J., Liebling, D., and Thompson, A. L. (2018). Multitasking with play write, a mobile microproductivity writing tool. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 411–422. 22, 48

[101] Irani, L. C. and Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 611–620. 7, 46, 49

[102] Iwata, T., Yamabe, T., and Nakajima, T. (2010). Towards a mobility enhanced user interface design for multi-task environments: An experimental study on cognitive workload measurement. In *2010 Sixth International Conference on Intelligent Environments*, pages 106–111. IEEE. 95, 170

[103] Jacques, J. T. and Kristensson, P. O. (2017). Design strategies for efficient access to mobile device users via amazon mechanical turk. In *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications*, pages 25–30. 47

[104] Jokela, T., Ojala, J., and Olsson, T. (2015). A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3903–3912. 8

[105] Jonnatan, L., Seaton, C. L., Rush, K. L., Li, E. P., and Hasan, K. (2022). Mobile device usage before and during the covid-19 pandemic among rural and urban adults. *International Journal of Environmental Research and Public Health*, 19(14):8231. 175

[106] Kaplan, T., Saito, S., Hara, K., and Bigham, J. P. (2018). Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*. 49, 51, 52

[107] Karlson, A. K., Iqbal, S. T., Meyers, B., Ramos, G., Lee, K., and Tang, J. C. (2010). Mobile taskflow in context: a screenshot study of smartphone usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2009–2018. 96

[108] Karlson, A. K., Meyers, B. R., Jacobs, A., Johns, P., and Kane, S. K. (2009). Working overtime: Patterns of smartphone and pc usage in the day of an information worker. In *International Conference on Pervasive Computing*, pages 398–405. Springer. 48, 96

[109] Kässi, O. and Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. *Technological forecasting and social change*, 137:241–248. 1

[110] Katsuragawa, K., Wallace, J. R., and Lank, E. (2016). In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 220–223. 75

[111] Kavin, B. P. and Ganapathy, S. (2021). A new digital signature algorithm for ensuring the data integrity in cloud using elliptic curves. *Int. Arab J. Inf. Technol.*, 18(2):180–190. 169

[112] Kay, J. (2008). Lifelong learner modeling for lifelong personalized pervasive learning. *IEEE Transactions on Learning Technologies*, 1(4):215–228. 165

[113] Khan, M. T., Hyun, M., Kanich, C., and Ur, B. (2018). Forgotten but not gone: Identifying the need for longitudinal data management in cloud storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12. 21

[114] Kim, Y.-H., Chou, D., Lee, B., Danilovich, M., Lazar, A., Conroy, D. E., Kacorri, H., and Choe, E. K. (2022). Mymove: Facilitating older adults to collect in-situ activity labels on a smartwatch with speech. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–21. 73

[115] Kinsella, B. and Mutchler, A. (2019). Smart speaker consumer adoption report. *Informe estadístico*. 10

[116] Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1301–1318. 2, 104

[117] Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P., Pina, P., and Singh, M. (2015). The global opportunity in online outsourcing. 1

[118] Kwon, J. (2022). *Machine Learning for AI-Augmented Design Space Exploration of Computer Systems*. Columbia University. 95

[119] Lago, P. P., Beruvides, M. G., Jian, J.-Y., Canto, A. M., Sandoval, A., and Taraban, R. (2007). Structuring group decision making in a web-based environment by using the nominal group technique. *Computers & Industrial Engineering*, 52(2):277–295. 17

[120] Lascau, L. (2023). *Time as Currency: A Value Sensitive Design Exploration of Crowdworkers' Temporal Flexibility*. PhD thesis, UCL (University College London). 2

[121] Lasecki, W. S., Kim, J., Rafter, N., Sen, O., Bigham, J. P., and Bernstein, M. S. (2015). Apparition: Crowdsourced user interfaces that come to life as you sketch them. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1925–1934. 21

[122] Laslo, Z., Keren, B., and Ilani, H. (2008). Minimizing task completion time with the execution set method. *European Journal of Operational Research*, 187(3):1513–1519. 21

[123] Law, E. and Ahn, L. v. (2011). Human computation. *Synthesis lectures on artificial intelligence and machine learning*, 5(3):1–121. 20

[124] Lazar, J., Allen, A., Kleinman, J., and Malarkey, C. (2007). What frustrates screen reader users on the web: A study of 100 blind users. *International Journal of human-computer interaction*, 22(3):247–269. 75

[125] Lee, J., Lee, C., and Kim, G. J. (2017). Vouch: multimodal touch-and-voice input for smart watches under difficult operating conditions. *Journal on Multimodal User Interfaces*, 11:289–299. 76

[126] Lee, S., Kyung, G., Lee, J., Moon, S. K., and Park, K. J. (2016). Grasp and index finger reach zone during one-handed smartphone rear interaction: effects of task type, phone width and hand length. *Ergonomics*, 59(11):1462–1472. 28

[127] Lehdonvirta, V. (2018). Flexibility in the gig economy: managing time on three online piecework platforms. *New Technology, Work and Employment*, 33(1):13–29. 51

[128] Li, J., Wang, Y., Yu, D., and Liu, C. (2021). Solvers' committed resources in crowdsourcing marketplace: do task design characteristics matter? *Behaviour & Information Technology*, pages 1–20. 22

[129] Li, L. and Lin, T. T. (2019). Smartphones at work: a qualitative exploration of psychological antecedents and impacts of work-related smartphone dependency. *International Journal of Qualitative Methods*, 18:1609406918822240. 9

[130] Liu, M., Zhang, F., and Huang, H. (2017). An adaptive classification strategy for reliable locomotion mode recognition. *Sensors*, 17(9):2020. 170

[131] Locke, P. R. (2005). *A tool for accessing corporate intranets to increase the productivity of mobile knowledge workers: An evaluation of a handheld wireless device*. PhD thesis, Capella University. 9

[132] Long, T. and Chilton, L. (2023). Ai and design opportunities for smart speakers. 73

[133] Lozovsky, A. (2023). The future of online security: How ai is revolutionizing captcha. 179

[134] Luckhaupt, S. E. and Tamers, S. L. (2021). Demographics of the u.s. workforce: A framework for considering implications for workplace safety and health. Accessed: 2024-05-16. 100

[135] Lv, Z., Poiesi, F., Dong, Q., Lloret, J., and Song, H. (2022). Deep learning for intelligent human–computer interaction. *Applied Sciences*, 12(22):11457. 95

[136] Maassen, O., Fritsch, S., Gantner, J., Deffge, S., Kunze, J., Marx, G., and Bickenbach, J. (2020). Future mobile device usage, requirements, and expectations of physicians in german university hospitals: web-based survey. *Journal of medical Internet research*, 22(12):e23955. 96

[137] Mao, J.-Y., Vredenburg, K., Smith, P. W., and Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, 48(3):105–109. 168

[138] Martin, F. and Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online learning*, 22(1):205–222. 165

[139] Martin, M. et al. (2021). Computer and internet use in the united states: 2018. *Suitland, MD: United States Census Bureau*. 9

[140] Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior research methods*, 44(1):1–23. 1

[141] McFarlane, D. C. and Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1):1–61. 95

[142] McGraw, I., Lee, C.-y., Hetherington, I. L., Seneff, S., and Glass, J. R. (2010). Collecting voices from the cloud. In *LREC*, pages 1576–1583. 21

[143] Mendel, T. and Toch, E. (2017). Susceptibility to social influence of privacy behaviors: Peer versus authoritative sources. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 581–593. 1

[144] Middleton, C. A. and Cukier, W. (2006). Is mobile email functional or dysfunctional? two perspectives on mobile email usage. *European Journal of Information Systems*, 15(3):252–260. 9

[145] Moore, D. R. (2006). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. 165

[146] Mullen, B., Johnson, C., and Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and applied social psychology*, 12(1):3–23. 18

[147] Müller, H., Gove, J. L., Webb, J. S., and Cheang, A. (2015). Understanding and comparing smartphone and tablet use: Insights from a large-scale diary study. In *Proceedings*

of the annual meeting of the australian special interest group for computer human interaction, pages 427–436. 9

[148] Murfin, M. (2013). Know your apps: an evidence-based approach to evaluation of mobile clinical applications. *The journal of physician assistant education: the official journal of the Physician Assistant Education Association*, 24(3):38–40. 10

[149] Murgia, M. (2019). Ai's new workforce: The data-labelling industry spreads globally. *Financial Times*. 1

[150] Myers, B., Hudson, S. E., and Pausch, R. (2000). Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):3–28. 167

[151] Nakatsu, R. T., Grossman, E. B., and Iacovou, C. L. (2014). A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science*, 40(6):823–834. 47

[152] Narula, P., Gutheim, P., Rolnitzky, D., Kulkarni, A., and Hartmann, B. (2011). Mobileworks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. *Human Computation*, 11(11):45. 11

[153] Nash, C., Jarrahi, M. H., and Sutherland, W. (2021). Nomadic work and location independence: The role of space in shaping the work of digital nomads. *Human Behavior and Emerging Technologies*, 3(2):271–282. 97

[154] Nasir, S. and Yurder, Y. (2015). Consumers' and physicians' perceptions about high tech wearable health products. *Procedia-Social and Behavioral Sciences*, 195:1261–1267. 10

[155] Nebeling, M. and Dey, A. K. (2016). Xdbrowser: user-defined cross-device web page designs. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5494–5505. 48

[156] Nebeling, M., Mintsi, T., Husmann, M., and Norrie, M. (2014). Interactive development of cross-device user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2793–2802. 48

[157] Nebeling, M., To, A., Guo, A., de Freitas, A. A., Teevan, J., Dow, S. P., and Bigham, J. P. (2016). Wearwrite: Crowd-assisted writing from smartwatches. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3834–3846. 8, 51, 78

[158] Neirotti, P., Raguseo, E., and Gastaldi, L. (2019). Designing flexible work practices for job satisfaction: the relation between job characteristics and work disaggregation in different types of work arrangements. *New technology, work and employment*, 34(2):116–138. 2, 12

[159] Newlands, G. and Lutz, C. (2021). Crowdwork and the mobile underclass: Barriers to participation in india and the united states. *new media & society*, 23(6):1341–1361. 3, 8, 11, 54, 71, 77, 99

[160] Nikulin, M. (1973). Chi-squared test for normality. In *Proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics*, volume 2, pages 119–122. 42

[161] Nouri, E., Fourney, A., Sim, R., and White, R. W. (2019). Supporting complex tasks using multiple devices. In *Proceedings of WSDM'19 Task Intelligence Workshop (TI@ WSDM19)*. 48

[162] Nouwens, M. and Nylandsted Klokmose, C. (2021). A survey of digital working conditions of danish knowledge workers. 97

[163] Oppenlaender, J., Opoku Asare, K., and Hosio, S. (2018). Campustracker: Assessing mobile workers' momentary willingness to work on paid crowdsourcing tasks. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 648–653. 8

[164] Oulasvirta, A., Rattenbury, T., Ma, L., and Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous computing*, 16:105–114. 164

[165] Oulasvirta, A. and Sumari, L. (2007). Mobile kits and laptop trays: managing multiple devices in mobile information work. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1127–1136. 9, 48, 96, 97

[166] Ozkale, A. and Koc, M. (2020). Investigating academicians' use of tablet pc from the perspectives of human computer interaction and technology acceptance model. *International Journal of Technology in Education and Science*, 4(1):37–52. 8, 9

[167] Pangakis, N., Wolken, S., and Fasching, N. (2023). Automated annotation with generative ai requires validation. *arXiv preprint arXiv:2306.00176*. 178

[168] Park, H. S. (2017). Technology convergence, open innovation, and dynamic economy. *Journal of Open Innovation: Technology, Market, and Complexity*, 3(4):24. 177

[169] Pearson, J., Robinson, S., Reitmaier, T., Jones, M., Ahire, S., Joshi, A., Sahoo, D., Maravi, N., and Bhikne, B. (2019). Streetwise: Smart speakers vs human help in public slum settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13. 10

[170] Peng, X., Gu, J., Tan, T. H., Sun, J., Yu, Y., Nuseibeh, B., and Zhao, W. (2016). Crowdservice: serving the individuals through mobile crowdsourcing and service composition. In *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering*, pages 214–219. 22

[171] Perrin, A. (2021). Digital divide persists even as americans with lower incomes make gains in tech adoption. Accessed: 2024-05-16. 100

[172] Pitt-Catsouphes, M. and Smyer, M. A. (2006). *One size doesn't fit all: Workplace flexibility.* Center on Aging & Work/Workplace Flexibility, Boston College. 175

[173] Piugie, Y. B. W., Di Manno, J., Rosenberger, C., and Charrier, C. (2021). How artificial intelligence can be used for behavioral identification? In *2021 International Conference on Cyberworlds (CW)*, pages 246–253. IEEE. 178

[174] Pizza, S., Brown, B., McMillan, D., and Lampinen, A. (2016). Smartwatch in vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5456–5469. 10

[175] Posch, L., Bleier, A., Lechner, C. M., Danner, D., Flöck, F., and Strohmaier, M. (2019). Measuring motivations of crowdworkers: The multidimensional crowdworker motivation scale. *ACM Transactions on Social Computing*, 2(2):1–34. 2

[176] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd. 15

[177] Qiu, L., Kanski, B., Doerksen, S., Winkels, R., Schmitz, K. H., and Abdullah, S. (2021). Nurse amie: Using smart speakers to provide supportive care intervention for women with metastatic breast cancer. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7. 10

[178] Raghuram, S. and Wiesenfeld, B. (2004). Work-nonwork conflict and job stress among virtual workers. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, 43(2-3):259–277. 2

[179] Rane, N., Choudhary, S., and Rane, J. (2023). Integrating chatgpt, bard, and leading-edge generative artificial intelligence in architectural design and engineering: applications, framework, and challenges. 178

[180] Reichenberger, I. (2018). Digital nomads–a quest for holistic freedom in work and leisure. *Annals of Leisure Research*, 21(3):364–380. 97

[181] Reyes, G., Zhang, D., Ghosh, S., Shah, P., Wu, J., Parnami, A., Bercik, B., Starner, T., Abowd, G. D., and Edwards, W. K. (2016). Whoosh: non-voice acoustics for low-cost, hands-free, and rapid input on smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pages 120–127. 73

[182] Rodrigues, J. G., Aguiar, A., and Barros, J. (2014). Sensemycity: Crowdsourcing an urban sensor. *arXiv preprint arXiv:1412.2070*. 21

[183] Rossitto, C. (2009). *Managing work at several places: Understanding nomadic practices in student groups*. PhD thesis, KTH. 7

[184] Roto, V. (2005). Browsing on mobile phones. *Nokia Research Center*, 10:2005. 22

[185] Ruoti, S., O'Neill, M., Zappala, D., and Seamons, K. (2016). User attitudes toward the inspection of encrypted traffic. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 131–146. 1

[186] Santos, J., Wauters, T., Volckaert, B., and De Turck, F. (2021). Towards low-latency service delivery in a continuum of virtual resources: State-of-the-art and research directions. *IEEE Communications Surveys & Tutorials*, 23(4):2557–2589. 94

[187] Savage, S., Chiang, C. W., Saito, S., Toxtli, C., and Bigham, J. (2020). Becoming the super turker: Increasing wages via a strategy from high earning workers. In *Proceedings of The Web Conference 2020*, pages 1241–1252. 2, 7, 14

[188] Schirra, S. and Bentley, F. R. (2015). "it's kind of like an extra screen for my phone" understanding everyday uses of consumer smart watches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2151–2156. 10

[189] Sefati, S. S. and Halunga, S. (2023). Ultra-reliability and low-latency communications on the internet of things based on 5g network: Literature review, classification, and future research view. *Transactions on Emerging Telecommunications Technologies*, 34(6):e4770. 94

[190] Sela, A. (2018). Can computers be fair: How automated and human-powered online dispute resolution affect procedural justice in mediation and arbitration. *Ohio St. J. on Disp. Resol.*, 33:91. 169

[191] Şerban, C. and Todericiu, I.-A. (2020). Alexa, what classes do i have today? the use of artificial intelligence via smart speakers in education. *Procedia Computer Science*, 176:2849–2857. 10

[192] Shao, Y.-F., Chang-Ogimoto, M., Pointner, R., Lin, Y.-C., Wu, C.-T., and Chen, M. (2016). Swipekey: a swipe-based keyboard design for smartwatches. In *Proceedings of the*

18th International Conference on Human-Computer Interaction with Mobile Devices and Services, pages 60–71. 75

[193] Sharma, H. K., Choudhury, T., Soni, R., and Sharma, S. (2023). Human computer interface (hci) controlled ai enabled system for optimization. In *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 552–557. IEEE. 180

[194] Silberman, M. S., Irani, L., and Ross, J. (2010). Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):39–43. 49

[195] Silver, H. and Goldscheider, F. (1994). Flexible work and housework: Work and family constraints on women's domestic labor. *Social Forces*, 72(4):1103–1119. 2

[196] Singh, G., Delamare, W., and Irani, P. (2018). D-swime: A design space for smartwatch interaction techniques supporting mobility and encumbrance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13. 75

[197] Singh, S. K., Rathore, S., and Park, J. H. (2020). Blockiotintelligence: A blockchain-enabled intelligent iot architecture with artificial intelligence. *Future Generation Computer Systems*, 110:721–743. 179

[198] Sivakumar, N., Kumar, K., Easwaran, B., and Tabassum, H. (2023). Design and analysis of human computer interaction using ai intelligence. In *2023 International Conference on Disruptive Technologies (ICDT)*, pages 195–198. IEEE. 178

[199] SmartSkill97 (2023). 5 crucial role of entrepreneurship in economic development. 177

[200] Soegaard, M. (2015). Interaction styles. *The glossary of human computer interaction*. 21, 23

[201] Soffer, T. and Yaron, E. (2017). Perceived learning and students' perceptions toward using tablets for learning: The mediating role of perceived engagement among high school students. *Journal of Educational Computing Research*, 55(7):951–973. 10

[202] Soykan, E. et al. (2015). Views of students', teachers' and parents' on the tablet computer usage in education. *Kıbrıslı Eğitim Bilimleri Dergisi*, 10(3):228–244. 10

[203] Srivastava, A., Kapania, S., Tuli, A., and Singh, P. (2021). Actionable ui design guidelines for smartphone applications inclusive of low-literate users. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–30. 72

[204] Strauss, A. and Corbin, J. M. (1997). *Grounded theory in practice*. Sage. 55, 79, 104

[205] Streitz, N. A., Tandler, P., Müller-Tomfelde, C., and Konomi, S. (2001). Roomware: Towards the next generation of human-computer interaction based on an integrated design of real and virtual worlds. *Human-computer interaction in the new millennium*, 553:578. 94

[206] Sun, H. and Zhang, P. (2006). The role of moderating factors in user technology acceptance. *International journal of human-computer studies*, 64(2):53–78. 167

[207] Swaminathan, S., Hara, K., and Bigham, J. P. (2017). The crowd work accessibility problem. In *Proceedings of the 14th International Web for All Conference*, pages 1–4. 33, 47

[208] Taivalsaari, A. and Mikkonen, T. (2017). The web as a software platform: Ten years later. In *International Conference on Web Information Systems and Technologies*, volume 2, pages 41–50. SCITEPRESS. 166

[209] TechMagic (2023). Ai anomaly detection: How it works and benefits for business. 180

[210] Terrenghi, L., Kronen, M., and Valle, C. (2005). Usability requirements for mobile service scenarios. *Human Computer Interaction*, pages 1–10. 95, 170

[211] Terzopoulos, G. and Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. *Informatics in Education*, 19(3):473–490. 8

[212] Toxtli, C., Richmond-Fuller, A., and Savage, S. (2020). Reputation agent: Prompting fair reviews in gig markets. In *Proceedings of The Web Conference 2020*, pages 1228–1240. 2, 14

[213] United Nations Economic Commission for Europe (2024). Employment and social protection. 176

[214] Uzor, S., Jacques, J. T., Dudley, J. J., and Kristensson, P. O. (2021). Investigating the accessibility of crowdwork tasks on mechanical turk. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14. 25, 46, 47

[215] Van Velsen, L., van der Geest, T., ter Hedde, M., and Derks, W. (2009). Requirements engineering for e-government services: A citizen-centric approach and case study. *Government Information Quarterly*, 26(3):477–486. 165

[216] Varanasi, R. A., Siddarth, D., Seshadri, V., Bali, K., and Vashistha, A. (2022). Feeling proud, feeling embarrassed: Experiences of low-income women with crowd work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–18. 180

[217] Vashistha, A., Sethi, P., and Anderson, R. (2017). Respeak: A voice-based, crowd-powered speech transcription system. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1855–1866. 11, 14, 51, 75, 78

[218] Vashistha, A., Sethi, P., and Anderson, R. (2018). Bspeak: An accessible voice-based crowdsourcing marketplace for low-income blind people. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13. 75

[219] Venkatesh, V. and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315. 167

[220] Venkatesh, V. and Ramesh, V. (2006). Web and wireless site usability: Understanding differences and modeling use. *MIS quarterly*, pages 181–206. 46

[221] Vilone, G. and Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76:89–106. 83

[222] Vogels, E. (2021). Digital divide persists even as americans with lower incomes make gains in tech adoption. *Pew Research Center*. 49

[223] Wanderer, J. P., Sandberg, W. S., and Ehrenfeld, J. M. (2011). Real-time alerts and reminders using information systems. *Anesthesiology clinics*, 29(3):389. 168

[224] Wang, J., Yang, X., Li, Z., et al. (2022). Cloud data integrity verification algorithm based on data mining and accounting informatization. *Scientific Programming*, 2022. 169

[225] Wang, L., Yu, Z., Han, Q., Yang, D., Pan, S., Yao, Y., and Zhang, D. (2020). Compact scheduling for task graph oriented mobile crowdsourcing. *IEEE Transactions on Mobile Computing*. 21

[226] Warrens, M. J. (2011). Cohen's kappa is a weighted average. *Statistical Methodology*, 8(6):473–484. 25

[227] Widdicks, K., Bates, O., Hazas, M., Friday, A., and Beresford, A. R. (2017). Demand around the clock: time use and data demand of mobile devices in everyday life. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 5361–5372. 9

[228] Williams, A. C., Kaur, H., Iqbal, S., White, R. W., Teevan, J., and Fourney, A. (2019a). Mercury: Empowering programmers' mobile work practices with microproductivity. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 81–94. 22, 48

[229] Williams, A. C., Mark, G., Milland, K., Lank, E., and Law, E. (2019b). The perpetual work life of crowdworkers: How tooling practices increase fragmentation in crowdwork. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–28. 2, 3, 6, 11, 14, 22, 25, 30, 48, 49, 52, 54, 77, 78, 99

[230] Woodcock, B., Middleton, A., and Nortcliffe, A. (2012). Considering the smartphone learner: an investigation into student interest in the use of personal technology to enhance their learning. *Student Engagement and Experience Journal*, 1(1):1–15. 10

[231] Wu, X., Huang, D., Sun, Y.-E., Bu, X., Xin, Y., and Huang, H. (2017). An efficient allocation mechanism for crowdsourcing tasks with minimum execution time. In *International Conference on Intelligent Computing*, pages 156–167. Springer. 21

[232] Xia, H., Wang, Y., Huang, Y., and Shah, A. (2017). " our privacy needs to be protected at all costs" crowd workers' privacy experiences on amazon mechanical turk. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–22. 104

[233] Xie, Y., Li, F., Wu, Y., and Wang, Y. (2021). Hearfit: Fitness monitoring on smart speakers via active acoustic sensing. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–10. IEEE. 10

[234] Yan, T., Marzilli, M., Holmes, R., Ganesan, D., and Corner, M. (2009). mcrowd: a platform for mobile crowdsourcing. In *Proceedings of the 7th ACM conference on embedded networked sensor systems*, pages 347–348. 11

[235] Yang, J. and Wigdor, D. (2014). Panelrama: enabling easy specification of cross-device web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2783–2792. 166

[236] Yin, M., Suri, S., and Gray, M. L. (2018). Running out of time: The impact and value of flexibility in on-demand crowdwork. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–11, New York, NY, USA. Association for Computing Machinery. 2, 51

[237] Yousaf, S., Haque, H. M. U., Atif, M., Hashmi, M. A., Khalid, A., and Vinh, P. C. (2023). A context-aware multi-agent reasoning based intelligent assistive formalism. *Internet of Things*, 23:100857. 180

[238] Zenker, S. and Hobert, S. (2019). Design and implementation of a collaborative smartwatch application supporting employees in industrial workflows. 97

[239] Zhang, D. and Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International journal of human-computer interaction*, 18(3):293–308. 15

[240] Zuccon, G., Leelanupab, T., Whiting, S., Yilmaz, E., Jose, J. M., and Azzopardi, L. (2013). Crowdsourcing interactions: using crowdsourcing for evaluating interactive information retrieval systems. *Information retrieval*, 16(2):267–305. 21, 22

[241] Zyskowski, K., Morris, M. R., Bigham, J. P., Gray, M. L., and Kane, S. K. (2015). Accessible crowdwork? understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1682–1693. 2, 46, 47

Appendix

A Study Instrument For Understanding Crowdworkers Practices and Desire in MTurk

Section 1: Demographics

Q1. What is your Worker ID? (This is used to ensure that you will be paid on MTurk.)

____

Q2. Please specify the range in which you fall:

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 or older

Q3. Please specify the gender to which you identify most:

- Male
- Female
- Non-Binary
- Prefer not to say

Q4. Please specify your highest level of education achieved:

- High School Diploma / GED Equivalent
- Trade School
- Associate's Degree
- Bachelor's Degree
- Advanced Degree, e.g., Master's / PhD / JD

Q5. How long have you worked on Amazon Mechanical Turk?

- Less than 0.5 Years
- Between 0.5 - 1 Years
- Between 1 - 2 Years
- Between 2 - 5 Years
- Longer than 5 Years

Q6. How many hours do you normally work on Amazon Mechanical Turk each week?

- Less than 1 hour per week
- 1-5 Hours per week
- 5-10 Hours per week
- 10-20 Hours per week
- 20-30 Hours per week
- 30-40 Hours per week
- 40+ Hours per week

Q7. How many HITs have you completed on Amazon Mechanical Turk?

____

Q8. What is your HIT approval rate on Amazon Mechanical Turk?

____

Q9. Briefly describe the primary tools, scripts, etc., that you use as a crowdworker (e.g., MTurk Suite, Turkinator, etc). If you do not use tools or scripts, type "None".

____

Section 2: Understanding HIT Completion

Q10. Please indicate how you currently use the following devices for completing HITs on MTurk.

Device	**Never**	**Rarely**	**Sometimes**	**Often**	**Always**	**N/A**
Desktop / Laptop	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smartphone, e.g., iPhone	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Tablet, e.g., iPad	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smart Watch, e.g., Apple Watch	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smart Speakers, e.g., Alexa	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Q11. Please report the frequency at which you do the following:

a. I currently use my smartphone to complete HITs on MTurk ______________.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am at my primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am within walking distance of your primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am away from my house.

Q12. Please report the frequency at which you do the following:

a. I currently use my tablet to complete HITs on MTurk ______________.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am at my primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am within walking distance of your primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am away from my house.

Q13. Please report the frequency at which you do the following:

a. I currently use my smart watch to complete HITs on MTurk ______________.

Never	Rarely	[ ]

While I am at my primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am within walking distance of your primary computer that I use for MTurk.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am away from my house.

Q14. Please report the frequency at which you do the following:

a. I currently use my smart speaker to complete HITs on MTurk ______________.

Never	Rarely	Sometimes	Often	Very Often
[ ]	[ ]	[ ]	[ ]	

While I am at my primary computer that I use for MTurk.

	Never	Rarely	Sometimes	Often	Very Often
	[ ]	[ ]	[ ]	[ ]	

While I am within walking distance of your primary computer that I use for MTurk.

	Never	Rarely	Sometimes	Often	Very Often
x [ ]	[ ]	[ ]	[ ]	[ ]	

While I am away from my house.

Q15. Please briefly describe what types of HITs you currently try to complete on your smartphone:

____

Q16. Please briefly describe what types of HITs you currently try to complete on your tablet:

____

Q17. Please briefly describe what types of HITs you currently try to complete on your smart speaker:

____

Q18. Please briefly describe what types of HITs you currently try to complete on your smart watch:

____

Q19. Consider the following types of HITs:

- Sentiment Analysis: HITs that ask you to rate the sentiment of text.
- Information Finding: HITs that ask you to find information on another website or search a paragraph of text.
- Audio Tagging: HITs that ask you to update meta-data fields in audio files.

- Speech Transcription: HITs that ask you to convert speech into a written document.
- Image Classification: HITs that ask you to classify images.
- Bounding Box: HITs that ask you to annotate a drawn rectangle over an image.

Based on your experience, please report the suitability of each type of device for completing each type of HIT on each device below. Please fill in each field below with a number between 1 to 5 where 1 indicates "Least Suitable" and 5 indicates "Most Suitable".

Based on your experience, please report the suitability of each type of device for completing each type of HIT on each device below. Please fill in each field below with a number between 1 to 5 where 1 indicates "Least Suitable" and 5 indicates "Most Suitable".

Task	**Workstation**	**Smartphone**	**Tablet**	**Smart Speaker**	**Smart Watch**
Sentiment Analysis	_____	_____	_____	_____	_____
Information Finding	_____	_____	_____	_____	_____
Audio Tagging	_____	_____	_____	_____	_____
Speech Transcription	_____	_____	_____	_____	_____
Image Classification	_____	_____	_____	_____	_____
Bounding Box	_____	_____	_____	_____	_____

Q20. For your MTurk work, what other types of HITs would you like to see better supported on the smartphone?

____

Q21. For your MTurk work, what other types of HITs would you like to see better supported on the tablet?

____

Q22. For your MTurk work, what other types of HITs would you like to see better supported on the smart speaker?

____

Q23. For your MTurk work, what other types of HITs would you like to see better supported on the smart watch?

____

Q24. From your experience, what motivates you to use (or not to use) a particular device to complete each of the above-mentioned types of HITs?

____

Section 3: Understanding HIT Management

Mobile device support for HIT management:

In this section, we inquire about the devices that you use to perform tasks related to managing HITs, such as finding HITs, reviewing HITs, and sharing HITs.

Q19. Please indicate how you currently use the following devices to find HITs, review HITs, or share HITs with other crowdworkers.

Device	**Never**	**Rarely**	**Sometimes**	**Often**	**Always**	**N/A**
Desktop / Laptop	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smartphone, e.g., iPhone	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Tablet, e.g., iPad	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smart Watch, e.g., Apple Watch	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Smart Speakers, e.g., Alexa	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Q20. Based on your experience, please report the effectiveness of each type of task related to managing HITs on each device below. Please fill in each field below with a number between 1 to 5 where 1 indicates "Not Effective at All" and 5 indicates "Very Effective".

Task	Workstation	Smartphone	Tablet	Smart Speaker	Smart Watch
Finding HITs	_____	_____	_____	_____	_____
Auto-Accepting HITs	_____	_____	_____	_____	_____
Creating Catchers/Watchers	_____	_____	_____	_____	_____
Listening to Catchers/Watchers	_____	_____	_____	_____	_____
Talking to Crowdworkers	_____	_____	_____	_____	_____
Talking to Requesters	_____	_____	_____	_____	_____

Q21. Which of these supporting tasks would you like to see better supported on smartphone?

- Finding HITs
- Auto-Accepting HITs
- Creating Catchers/ Watchers
- Listening to Catchers / Watchers
- Talking to Crowdworkers/ Other MTurk Workers
- Talking to Requesters
- None of the above

Q22. Which of these supporting tasks would you like to see better supported on tablet?

- Finding HITs
- Auto-Accepting HITs
- Creating Catchers/ Watchers
- Listening to Catchers / Watchers

- Talking to Crowdworkers/ Other MTurk Workers
- Talking to Requesters
- None of the above

Q23. Which of these supporting tasks would you like to see better supported on your smart watch?

- Finding HITs
- Auto-Accepting HITs
- Creating Catchers/ Watchers
- Listening to Catchers / Watchers
- Talking to Crowdworkers/ Other MTurk Workers
- Talking to Requesters
- None of the above

Q24. Which of these supporting tasks would you like to see better supported on your smart speaker?

- Finding HITs
- Auto-Accepting HITs
- Creating Catchers/ Watchers
- Listening to Catchers / Watchers
- Talking to Crowdworkers/ Other MTurk Workers
- Talking to Requesters
- None of the above

Q25. From your experience, what motivates you to use (or not to use) a particular device for each of these tasks?

____

Section 4: Broken Desktop

Consider the following scenario:

Suppose the desktop / workstation computers that you normally use to complete work on Mechanical Turk suddenly break. With this in mind, please answer the following questions:

Q23. Please state your agreement with the following statement:

	Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
My computer being broken is disruptive to my MTurk work.	[ ]	[ ]	[ ]	[ ]	[ ]

Q24. In the context of your broken workstation, pretend that you must continue working. Please select the "alternative device" which you would resort to using to complete HITs on Amazon Mechanical Turk instead:

- Smartphone
- Tablet
- Smart Speaker
- Smart Watch

Q25. Why did you select this as an alternative device?

____

Q26. As your computer is unavailable, pretend that you've now decided to try to work on Mechanical Turk with the "alternative device" you selected above. With this in mind, please answer the extent that you agree with the following statements.

	Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
I have access to the appropriate tools (e.g., MTurk Suite, HIT Scraper, etc) to work effectively from my alternative device.	[ ]	[ ]	[ ]	[ ]	[ ]
The ability to find and accept HITs from my alternative device would improve my productivity.	[ ]	[ ]	[ ]	[ ]	[ ]
The ability to perform HITs from my alternative device would improve my productivity.	[ ]	[ ]	[ ]	[ ]	[ ]
I will be just as productive working from my alternative device as I would have been working from my workstation computer.	[ ]	[ ]	[ ]	[ ]	[ ]
After my computer is fixed, I will be able to continue the work on my workstation that I was doing on my alternative device.	[ ]	[ ]	[ ]	[ ]	[ ]

Section 5: The Magic Wand

Making MTurk Better:

You have been given a Magic Wand that allows you to change whatever you'd like to change about work on Amazon Mechanical Turk to work on the platform how you want to work.

Q25. How would you use the magic wand to make managing and performing HITs better for a desktop/laptop? What would you change? Why?

____

Q26. How would you use the magic wand to make managing and performing HITs better for a smartphone? What would you change? Why?

____

Q27. How would you use the magic wand to make managing and performing HITs better for a tablet? What would you change? Why?

____

Q28. How would you use the magic wand to make managing and performing HITs better for a smart speaker? What would you change? Why?

____

Q29. How would you use the magic wand to make managing and performing HITs better for a smart watch? What would you change? Why?

____

B Study Instrument For Understanding Crowdworkers Practices and Desire in Prolific

General Demographics

Q1. What is the highest level of education you have completed?

- Less than high school
- High school graduate
- Some college, no degree
- 2 year college degree
- 4 year college degree
- Graduate degree (MS, PhD, MD, JD, etc.)

Q2. Using your own words, please describe your ethnicity or cultural background.

____

Technological Demographics

Q1. Which of the following devices do you own? Select all that apply.

- Desktop or laptop
- Tablet
- Smartphone
- Smartwatch
- Smart speaker

Q2. How often do you use the following devices (not just for crowdwork)?

Device	**Always**	**Very Often**	**Sometimes**	**Rarely**	**Never**
Desktop or laptop	[ ]	[ ]	[ ]	[ ]	[ ]
Tablet	[ ]	[ ]	[ ]	[ ]	[ ]
Smartphone	[ ]	[ ]	[ ]	[ ]	[ ]
Smartwatch	[ ]	[ ]	[ ]	[ ]	[ ]
Smart speaker	[ ]	[ ]	[ ]	[ ]	[ ]

Crowdwork Demographics

Q1. Please indicate how much you agree with the following statements.

Statement	Strongly Agree	Somewhat Agree	Neither Agree norDisagree	Somewhat Disagree	Strongly Disagree
I am eager to try out new devices and software	[ ]	[ ]	[ ]	[ ]	[ ]
I prefer to use devices and software I'm already comfortable with	[ ]	[ ]	[ ]	[ ]	[ ]

Q2. Which of the following platforms do you currently use as a crowdworker? Select all that apply.

- Prolific
- Amazon Mechanical Turk
- HeroX
- Upwork
- InnoCentive
- Topcoder
- CrowdFlower
- Other platform (please specify)

Q3. Which of the following tools and scripts do you use for crowdwork? Select all that apply.

- Turk Guru
- Turker View

- Panda Crazy
- Stax
- Quebicle
- MTurk Suite
- HIT Forker
- Otto
- Other tool (please specify)
- I do not use any tool

Crowdwork Activity

Q1. Roughly, how many years have you been doing crowdwork? If 10 or more years, select 10.

Years: [ ]0 [ ]1 [ ]2 [ ]3 [ ]4 [ ]5 [ ]6 [ ]7 [ ]8 [ ]9 [ ]10

Q2. Roughly, how many hours a week do you spend completing crowdwork? If 50 or more hours, select 50.

Hours: [ ]0 [ ]5 [ ]10 [ ]15 [ ]20 [ ]25 [ ]30 [ ]35 [ ]40 [ ]45 [ ]50

Q3. Roughly, how many hours a week do you spend managing crowdwork? If 10 or more hours, select 10.

Hours: [ ]0 [ ]1 [ ]2 [ ]3 [ ]4 [ ]5 [ ]6 [ ]7 [ ]8 [ ]9 [ ]10

Financial Demographics

Q1. Roughly, what percentage of your total income comes from crowdwork?

Percent: [ ]0 [ ]25 [ ]50 [ ]75 [ ]100

Q2. Which of the following best describes how your income compares to others where you live?

- Far below average
- Below average
- Average
- Above average
- Far above average

Q3. How many dependents are you responsible for financially supporting (children, parents, etc.)? If 4 or more, select 4.

Dependents: [ ]0 [ ]1 [ ]2 [ ]3 [ ]4

Primary Profession

Q1. Outside of crowdwork, which of the following best describes your primary profession?

- Healthcare
- Education
- Information and communication technology
- Management
- Agricultural, forestry, or fishery
- Service and sales
- Craftsman or tradesman
- Other (please specify)

Q2. Outside of crowdwork, is your primary profession completed in person or remotely?

- It is fully in person
- It is fully remote
- It is a hybrid position, with both remote and on-site work requirements

Device Usage

Q1. In regards to crowdwork, how often do you use your *device* to complete and manage crowdwork?

	Always	**Very Often**	**Sometimes**	**Rarely**	**Never**
Complete crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]
Manage crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]

Q2. When choosing whether to use your *device* to complete or manage crowdwork, how important are the following factors in that decision?

Factor	**Very Important**	**Somewhat Important**	**It is Not Important**
Whether I am at or away from home	[ ]	[ ]	[ ]
The time of day	[ ]	[ ]	[ ]
A lack of crowdwork tools that work well on a *device*	[ ]	[ ]	[ ]
It is difficult to effectively manage my crowdwork on a *device*	[ ]	[ ]	[ ]
A lot of crowdwork explicitly prohibits the use of a *device*	[ ]	[ ]	[ ]
A lot of crowdwork is not well formatted for a *device*	[ ]	[ ]	[ ]
Other challenges (please specify)	[ ]	[ ]	[ ]

Q3. Now, imagine an ideal world where improved, high-quality tooling for completing and managing crowdwork is available for a *device*. In that world, how often would you use a *device* to complete and manage crowdwork?

	Always	**Very Often**	**Sometimes**	**Rarely**	**Never**
Complete crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]
Manage crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]

Crowdworkers' Opinions

Q1. Considering the ideal world again, how interested would you be in using a *device* to complete the following types of crowdwork?

Task Type	**Very Interested**	**Somewhat Interested**	**Not Interested**
Surveys	[ ]	[ ]	[ ]
Testing software or systems	[ ]	[ ]	[ ]
Information finding	[ ]	[ ]	[ ]
Classification, categorization, labeling, tagging, ranking, etc.	[ ]	[ ]	[ ]
Transcription or translation	[ ]	[ ]	[ ]

Q2. Considering the ideal world again, how interested would you be in using a *device* to perform the following management tasks?

Task Type	**Very Interested**	**Somewhat Interested**	**Not Interested**
Communicating with other fellow crowdworkers	[ ]	[ ]	[ ]
Receiving notifications (available crowdwork, rejected crowdwork, etc.)	[ ]	[ ]	[ ]
Finding and accepting suitable crowdwork	[ ]	[ ]	[ ]

Communicating with requesters or crowdsourcing platforms	[ ]	[ ]	[ ]

Q1. In regards to completing and managing crowdwork, how important are the following items to you?

	Extremely Important	**Very Important**	**Moderately Important**	**Slightly Important**	**Not at all Important**
The ability to do crowdwork at the place of my choosing	[ ]	[ ]	[ ]	[ ]	[ ]
Being provided with ample time to complete crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]
Being able to choose between a variety of types of crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]
The ability to do crowdwork using the device of my choosing	[ ]	[ ]	[ ]	[ ]	[ ]

The ability to do crowdwork at the time of my choosing	[ ]	[ ]	[ ]	[ ]	[ ]

Q2. When deciding which crowdwork platforms you will use, how important are each of the following items?

	Extremely Important	**Very Important**	**Moderately Important**	**Slightly Important**	**Not at all Important**
The platform supports effective communication between crowdworkers and requesters	[ ]	[ ]	[ ]	[ ]	[ ]
The platform supports completing crowdwork on non-desktop/laptop devices	[ ]	[ ]	[ ]	[ ]	[ ]

The platform supports managing crowdwork on non-desktop/laptop devices	[ ]	[ ]	[ ]	[ ]	[ ]
The platform provides sufficient features that crowdworkers don't feel the need to use other crowdwork tools/scripts	[ ]	[ ]	[ ]	[ ]	[ ]
The platform promotes the fair treatment of crowdworkers	[ ]	[ ]	[ ]	[ ]	[ ]
The platform promotes the fair treatment of requesters	[ ]	[ ]	[ ]	[ ]	[ ]

The platform enforces a minimum, fair compensation for crowdworkers	[ ]	[ ]	[ ]	[ ]	[ ]
The platform prevents bots from stealing crowdwork from legitimate crowdworkers	[ ]	[ ]	[ ]	[ ]	[ ]

Q3. When deciding which crowdwork platforms you will use, how important are each of the following items?

	Extremely Important	**Very Important**	**Moderately Important**	**Slightly Important**	**Not at all Important**
The platform makes it easy to find crowdwork	[ ]	[ ]	[ ]	[ ]	[ ]

The platform provides crowdwork that allows crowdworkers to make an impact in the world	[ ]	[ ]	[ ]	[ ]	[ ]
The platform has a strong community of crowdworkers	[ ]	[ ]	[ ]	[ ]	[ ]
The platform provides strong protections for crowdworkers' privacy and personal data	[ ]	[ ]	[ ]	[ ]	[ ]

Vita

Senjuti hails from Bardhaman, India, embarked on her academic journey in Computer Engineering, earning her Bachelor's degree from Meghnad Saha Institute of Technology, India. Following her undergraduate studies, she furthered her research experience as a research intern at Jadavpur University, India. Senjuti's quest for advanced knowledge led her to the University of Memphis, where she pursued and completed her Master's degree. Her master's thesis, titled "Enabling Efficient and Privacy-Preserving Task Matching," exemplifies her dedication to enhancing technology while prioritizing user privacy. Building upon her academic and research foundation, Senjuti then joined the University of Tennessee, Knoxville (UTK), to pursue her Ph.D. in Computer Science, marking the next chapter in her academic and professional development.