Why Automate This? Exploring Correlations Between Desire for Robotic Automation, Invested Time and Well-Being

Ruchira Ray*
University of Texas at Austin
ruchiraray@utexas.edu

Leona Pang
University of Texas at Austin
leonapang@utexas.edu

Sanjana Srivastava
Stanford University
sanjana2@stanford.edu

Li Fei-Fei
Stanford University
feifeili@stanford.edu

Samantha Shorey
University of Pittsburgh
samshorey@pitt.edu

Roberto Martín-Martín
University of Texas at Austin
robertomm@utexas.edu

Abstract

Understanding the motivations underlying the human inclination to automate tasks is vital for developing robots that fit seamlessly into daily life. Accordingly, we ask: are individuals more inclined to automate activities based on the time they consume or the feelings experienced while performing them? This study explores these preferences and whether they vary across social groups, specifically gender category and income level. Leveraging data from the BEHAVIOR-1K dataset, the American Time-Use Survey, and the American Time-Use Survey Well-Being Module, we investigate the relationship between the desire for robot automation, time spent, and associated feelings: Happiness, Meaningfulness, Sadness, Painfulness, Stressfulness, or Tiredness. Our key findings show that, despite common assumptions, time spent on activities does not strongly predict automation preferences; instead, happiness and pain are the strongest indicators. We also identify differences by gender and economic level: Women prefer to automate stressful activities, whereas men prefer to automate those that make them unhappy; mid-income individuals prioritize automating less enjoyable and meaningful activities, while low and high-income show no significant correlations. We hope our research helps motivate the design of robots that align with user priorities, moving domestic robotics toward more socially relevant solutions. All data and an interactive tool are publicly available at `https://robin-lab.cs.utexas.edu/why-automate-this/`.

1 Introduction

Robots offer a future in which people are alleviated from the labor of physical activities that they routinely perform, but would ideally prefer not to do. Unlike workplace automation, which is often driven by economic factors, the adoption of technology at home depends on consumer choices. To align household robotics research with user needs, Li et al. [1] established BEHAVIOR-1K, a benchmark that includes a ranked list of the activities people want robots to perform for them. The benchmark reveals *which* tasks people most want automated, not *why*: Are these tasks time-consuming or do they evoke unpleasant feelings (Fig. 1)?

As an initial step towards understanding what motivates automation in the home, we analyze information from three publicly available datasets: 1) the BEHAVIOR-1K (B1K) survey, reporting on the

*Now at the University of Edinburgh.

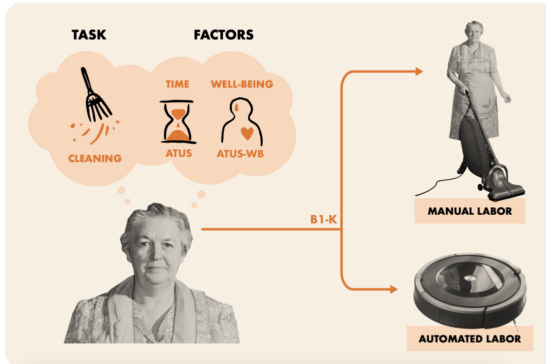


Figure 1: **What drives the desire for robotic automation?** Do people want to perform everyday household tasks themselves, or do they want a robot to do it? What motivates the decision to delegate tasks: a desire to save time or to avoid unpleasant experiences? To answer this question, we perform a statistical analysis with three data sources, B1K [1], ATUS [2] and ATUS-WB [3].(Figure by Helen Tseng. Original image sources: [4, 5])

desire for robotic automation based on thousands of individual surveys, 2) the American Time-Use Survey (ATUS [2]), a nationally representative survey conducted by the US Department of Labor, reporting on the time Americans spend on different everyday activities, and 3) the ATUS Wellbeing Module (ATUS-WB [3]), an additional dataset reporting on the feelings Americans experience when performing ATUS activities. Our contrastive analysis addresses the following research questions:

- **RQ1-** Does the average time spent (T) on an activity predict the desire for robotic automation (DA)?
- **RQ2-** Which feelings experienced while performing an activity are the strongest predictors of the desire for robotic automation? Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), or Tiredness (Z).

We also investigate differences between gender and income groups, as household robots will automate tasks more likely to be performed by women due to social divisions of labor [6]. In the U.S., women are primarily responsible for the category of activities ATUS defines as "housework" (laundry, sweeping the floor, putting away groceries) and spend twice as much time on these chores as men, who more often do other activities like yard work [2]. These delineations reflect long-standing cultural connotations of domestic tasks as "masculine" or "feminine" [7] that are evident in how housework is divided in heterosexual partnerships[8]. American mothers spend roughly the same amount of time on housework regardless of their economic strata [9], however in areas with high-income inequality, differences in time spent among high-earning and low-earning women are more pronounced [10].

We therefore seek to understand if or how these differing experiences affect what activities people prioritize for automation:

- **RQ3-** Do gender-based differences – in terms of the average time spent on an activity or the feelings associated with performing an activity – yield differences in desire for robotic automation?
- **RQ4-** Do income-based differences – in terms of the average time spent on an activity or the feelings associated with performing an activity – yield differences in desire for robotic automation?

Our study reveals that, while household technologies are often marketed as "time-saving devices" [11], there is no correlation between time spent on an activity and the desire to automate it. Instead, lower happiness during an activity strongly predicts a higher desire for robotic automation, as does painfulness; sadness and tiredness do not. Psychological theories of affective regulation suggest that individuals are motivated to seek positive experiences and avoid negative ones [12, 13, 14]. We therefore frame the desire for robotic automation as a form of affective regulation: people may wish to offload activities that are unpleasant, not merely time-consuming. Examining the data by gender, we find that men prioritize automating activities they are less happy doing, while women prioritize automating activities that are linked to stress. Both men and women prioritize automating activities that they spend comparatively less time on than their counterparts. Further, the mid-income group prioritizes automating activities that are experienced as less meaningful.

Additionally, we open-source all our data, including the B1K survey results, along with an interactive tool for data processing, statistical analysis, and visualization, available on a public website[2]. ATUS is a rich source of critical demographic data, yet it is not easy to parse; we hope that our tool lowers the barrier for others to replicate our results and further use it to guide future robotics research.

Statement of Contributions: This paper presents the first data-driven investigation into why people wish to automate household tasks, analyzing how time use and emotional experience relate to automation preferences across three large datasets: BEHAVIOR-1K, ATUS, and ATUS-WB. Building on prior work that identified what people want robots to do, we provide quantitative analysis exploring the emotional and demographic trends shaping automation preferences, offering insights for designing robots that align with user priorities. All data, code, and an interactive tool for dataset exploration and visualization are made publicly available.

2 Literature Review

Automating Household Activities: In the absence of artificial general intelligence, the specific task a robot accomplishes is a fundamental design decision – shaping functionality, behavior, and the parameters of plausible events [15]. While routine household activities may seem simple, they are technically complex and require a detailed understanding of task steps. Cakmak & Takayama's aggregation of digital "chore-lists" indicates that cleaning tasks make up nearly half of all domestic chores [16]. Cleaning tasks are united by the need to apply a tool to a stationary surface, yet each require different tool-use skills. HRI researchers have taken various approaches to collecting the information to support automation, including crowd-sourcing human explanations of task procedures [17, 18] and end-user demonstration [19].

The survey portion of the benchmark BEHAVIOR-1K [1] finds that participants' highest priorities for automation are nearly all classified as *housework* by the ATUS coding rules, with interior cleaning tasks (such as mopping the floor and scrubbing bathroom fixtures) at the top. This aligns with prior studies showing cleaning is the most common task people envision for robots [20, 21]. People also tend to prefer robots for hands-on, physical tasks [22]. Importantly, these priorities reflect not just time spent on tasks but also their unpleasantness, difficulty, level of dirt, or danger [23, 24, 25].

Household Activities as Household Experiences: Bell and Kay argue that designing technologies for the home requires pivoting from the dominant, commercial view where household activities are "simple, beginning-to-end processes" [26] and instead accounting for experience. Principles for feminist HRI also calls for moving beyond metrics of efficiency (which drive the development of robots in industrial contexts) towards users' bodily sensations and emotions [27]. We take up these calls by examining how experiences with household activities shape the desire for robotic automation.

Time Used for Household Activities: Decades of studies utilizing time use data, such as ATUS, show that women spend significantly more time on household activities than men [28, 29]. Further the type of household activities performed by women are often more time-consuming and repetitive, reinforcing the constant demand on women's time [30]. Women are also more likely to engage in multiple household tasks simultaneously, adding to their cognitive and physical load [31]. Despite women's persistent responsibility for housework across income levels [32], lower-income households face greater burdens due to limited resources to outsource domestic labor and increased time demands, leading to heightened stress for women and men [33, 34].

Impact of Household Robots on Time Use and Well-Being: The division of household labor raises important questions about whether the desire for robotic automation may be related to the social assignment of domestic activities. Automation is expected to significantly reduce the time spent on unpaid domestic work [35]. For instance, robotic vacuum cleaners have been shown to decrease time spent on cleaning, [36] fulfilling the commonly espoused goal of easing domestic burdens. Accordingly, international time use data suggests automation of household chores is likely to have a greater impact on women [35]. Tasks traditionally considered masculine (such as lawn maintenance) see less demand for automation and existing robotic solutions are less frequently adopted [37].

As Lee and Šabanović [38] observe, domestic technologies are shaped by social dynamics, not just functional needs. HRI studies demonstrate that the use of household robots is shaped by gendered, cultural norms around domestic responsibility *and* technical expertise [38, 39, 40, 41]. Though

[2] `https://robin-lab.cs.utexas.edu/why-automate-this/`

women traditionally dedicate more time to household tasks such as vacuuming, researchers find that men and children began taking on cleaning roles when robotic vacuums were introduced [41, 39]. This aligns with broader patterns in which men more often configure digital household technologies [42]. By shifting household dynamics and automating certain tasks, robots may help rebalance aspects of domestic responsibility, the burden of which has social and psychological impacts[3] [43].

Beyond saving time, scholars of the future of work stress the importance of preserving meaningful tasks that provide fulfillment and purpose [44, 45]. Within the domestic context, this opens questions about which daily activities bring satisfaction. Together, these perspectives highlight the need for HRI to better understand the desire for robotic automation: what opportunities for well-being arise from automating (or not automating) specific domestic tasks?

3 Methods

To address our research questions (RQ1-RQ4), we conduct an exploratory, statistical analysis of three publicly available datasets: BEHAVIOR-1K (B1K) [1], the American Time Use Survey (ATUS) [2], and the ATUS Well-being Module (ATUS-WB) [3]. B1K contains information about the desire for robotic automation; ATUS contains information about the time spent on activities; and ATUS-WB contains information about the feelings experienced during different activities (Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), and Tiredness (Z)).

3.1 Hypotheses

Our analysis aims to test working hypotheses about the desire for robotic automation (DA) and its correlations with time spent (T) and various well-being metrics:

(H1) *There is a positive correlation between the desire for robotic automation and the time people spend on a task.* We hypothesize that this general relationship T-DA remains invariant among social subgroups based on gender (G) and income level (I). However, the specific activities that each subgroup spends more/less time on varies: **(H1b)** *A relative increment in time spent by a social subgroup on a specific activity (compared to the general population and other subgroups) is directly related to a relative increment in the desire for robotic automation for that activity (also compared to the general population and other subgroups).*

(H2) *There is a correlation between the desire for robotic automation of an activity and the feelings experienced when performing this activity.* This correlation will be **(H2.1) negative** with respect to **happiness**, **(H2.2) negative** with respect to **meaningfulness**, **(H2.3) positive** with respect to **painfulness**, **(H2.4) positive** with respect to **sadness**, **(H2.5) positive** with respect to **stressfulness**, and **(H2.6) positive** with respect to **tiredness**. In addition to the general trends observed in the entire population, we hypothesize that *if a social subgroup experiences difference in feelings from an activity (compared to the general population and other subgroups), their desire to automate that activity will correspondingly shift.* The change will be **(H2.1b) negative** with respect to **happiness**, **(H2.2b) negative** with respect to **meaningfulness**, **(H2.3b) positive** with respect to **painfulness**, **(H2.4b) positive** with respect to **sadness**, **(H2.5b) positive** with respect to **stressfulness**, and **(H2.6b) positive** with respect to **tiredness**.

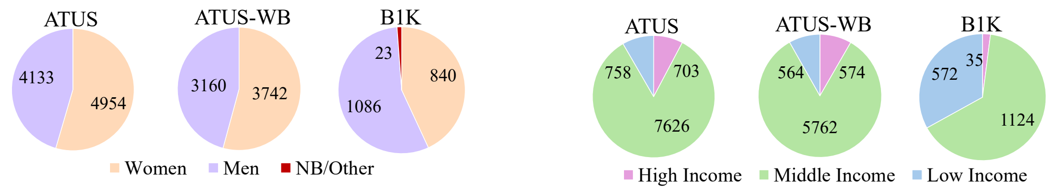
3.2 Research Design

3.2.1 Datasets

The BEHAVIOR-1K dataset (B1K) [1] comprises responses from 1,949 participants collected via Amazon Mechanical Turk. Participants were asked, "*On a scale of 1 to 10, rate how much you want a robot to do this activity for you?*" Responses were recorded on an independent Likert scale, with 1 representing "less beneficial" and 10 representing "most beneficial." Each participant rated 50 different tasks from approximately 2,000 tasks sourced from time-use surveys and WikiHow articles. Participant demographics are summarized in Fig. 2.

The American Time Use Survey (ATUS) [2] captures how Americans spend their time daily, logging activities by minute intervals with details such as activity type, duration, and demographic information

[3]However, these gains in equality (in terms of time) may come at the cost of upholding stereotypes around technical expertise.



(a) Gender distribution

(b) Income distribution

Figure 2: **Dataset Demographic Distribution:** Distribution of gender (left) and income levels (right) across B1K, ATUS, and ATUS-WB datasets; NB=Non-binary; ATUS and ATUS-WB have almost identical demographics; B1K is slightly different but shows similar trends (e.g., most participants are of middle-income class).

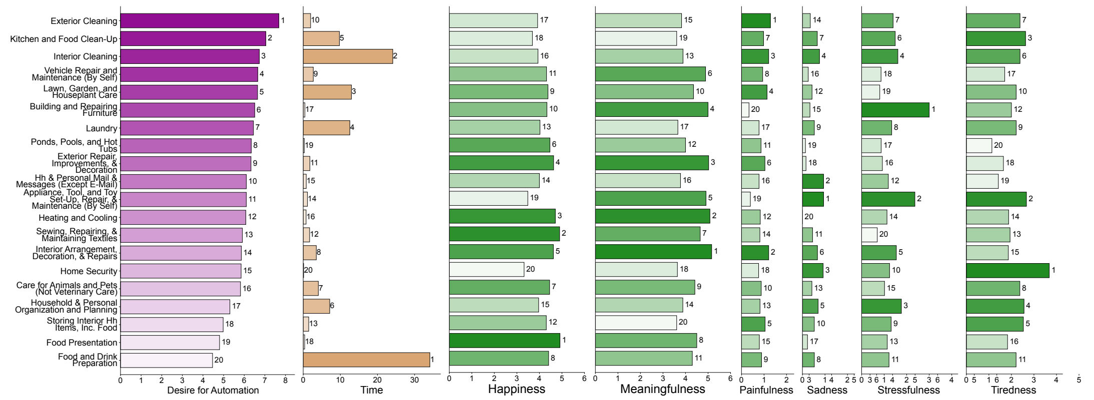


Figure 3: **Desire for Robotic Automation (DA), Time Spent (T), Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), and Tiredness (Z) for the General Population (GP) in the subset of *Household Activities*.** All ranks are ordered by DA. Darker color indicates higher rank; numbers on each bar indicate rank position. Visually, no clear trends are observable, except for a slight negative correlation between DA and H and a positive one between DA and P.

(e.g., sex, income, number of household members). This survey has been performed annually by the Bureau of Labor Statistics (BLS) since 2003.

The American Time Use Well-Being Module contains information related to how people felt during selected activities. The ATUS Well-Being Module is administered to a subset of ATUS respondents. Selected respondents are asked to rate how they felt during three activities they completed the previous day. Ratings are on a scale of 0 to 6 across six feelings: Happy, Meaningful, Pain, Sad, Stress and Tired. Our study used all 6 of the available factors from the ATUS-WB module. These were selected by the BLS and are not further defined in the documentation. The BLS let respondents interpret questions by their "common understanding" of the terms.

B1K offers detailed task data (e.g., *changing sheets*), while ATUS and ATUS-WB present higher-level activity categories (e.g., *interior cleaning*). To enable comparative analysis across the datasets, B1K tasks were manually coded using the extensive category definitions provided by ATUS for analysis of time-use diaries. We excluded *work, main job* across all datasets, as our analysis does not contend with automation of wage labor.

ATUS includes the category code *Household Activities (20000)* for tasks like cleaning, laundry, food preparation, yard care, and vehicle repair. For visual clarity, we use this subset in our main plots, while the full analysis, dataset curation, and alignment details are provided in the Appendix 7 in the extended version of this paper [46].

3.2.2 Dependent Variables

Our dependent variables are Desire for Robotic Automation (DA), Average Time Spent (T), and activity feelings: Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), and Tiredness (Z):

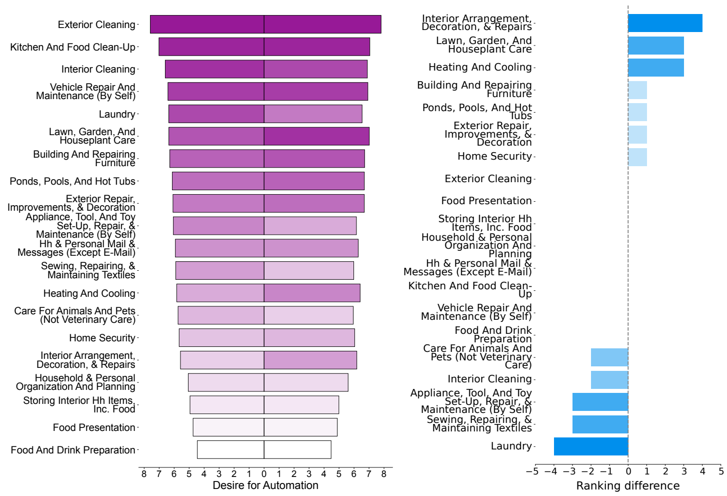


Figure 4: **Comparison of The Desire for Robotic Automation of Household Activities (Men vs. Women)**: (Left) Household activities ranked by desire for robotic automation from men on the left and corresponding DA for women on the right, a darker color indicates higher rank; (Right) $RankPosMen - RankPosWomen$: Ordered differences in ranking positions between men and women, a darker color indicates larger absolute difference. A positive value for activity indicates that the activity ranks further away from the top for men than for women, corresponding to $RankPosMen > RankPosWomen$.

Desire for Robotic Automation (DA): The DA score per activity was then computed by averaging the total scores of all tasks within each activity.

Average Time Spent on Activities (T): The average time spent was derived from the ATUS dataset, focusing on activities shared with B1K while excluding non-automatable ones like sleeping. Average time per activity was calculated by dividing the total recorded time by the number of entries.

There are two approaches for determining average time spent on activities. One averages over all participants, capturing how much time people spend on an activity daily— this provides a true population-level average, which we utilize in our research. The other approach averages only among those who did the activity, reflecting task duration but not overall prevalence. For example, ATUS indicates that the average time spent by the entire population on *Golfing* is approximately 0.02 hours, accounting for both participants and non-participants. Alternatively, for those who do golf, the average time spent is about 2.97 hours per day.

Activity Well-Being Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), Tiredness (Z) Score: The mean well-being score associated with each activity is calculated using the ATUS-WB Module. Consistency in activity-level comparison was maintained by selecting a subset of ATUS-WB data that corresponded to the activities included in our time calculation. The scores for an activity are computed by averaging the total sum of each metric by the number of its occurrences within the module.

3.2.3 Independent Variables

Our independent variables consist of subgroups that represent two demographic categories: gender subgroups (men and women) and income levels (low, middle, and high-income levels). These help us to isolate and analyze any distinct preferences and correlations within each demographic group.

Gender: The B1K [1] collected gender data, with participants identifying as women (43.41%), men (55.50%), non-binary (0.83%), or "other" (0.26%) ATUS adheres to a binary classification of sex, recognizing only male (53.8%) and female (46.2%). This binary classification does not represent the full spectrum of gender identities, potentially resulting in mismeasurement and misrepresentation of participants. In this paper, we analyze both the "male/female" response as reported in ATUS and the "man/woman" response as reported in B1K surveys as the independent variable "gender category:" men/women. Though these responses may reflect two different aspects of one's identity (biological sex vs. gender), our decision was driven by the need for comparability across the three datasets. We discuss the variable *Gender* further in our limitations section (Sec. 5) and in Appendix 7.6 [46].

Income: We divided ATUS and B1K participants into high, low, and mid-income levels based on household size, the poverty line, and income data from 2021. The upper boundary for the low-income category was taken from historical poverty thresholds from the U.S. Census Bureau [47]. The lower boundary for the high-income category was informed by the Pew Research Center's analysis [48]. The income distribution of BEHAVIOR 1K shows 64.9% mid-income, 33.0% low-income, and 2.0% high-income groups. Further explanation can be found in Appendix 7.3 [46] and our website.

3.3 Measures

To investigate the motivating factors behind the desire for robotic automation, we compare the ranked lists of activities based on the Desire for Robotic Automation to ranked lists based on time spent, and the ATUS-WB variables: Happiness, Meaningfulness, Painfulness, Sadness, Stressfulness and Tiredness (details on how activity-level rankings are constructed are provided in Appendix 7.4) [46]. Fig. 3 depicts the ranked activities for the entire population using the order of the Desire for Robotic Automation for the other variables. We use ranked lists since the absolute values for each variable are non-comparable (desire for robotic automation score vs. time [minutes] vs. well-being scores). To compare the ranked lists, we used non-parametric correlation tests: Spearman's rho and Kendall's tau. We performed these analyses on the general population and the subgroups. For all analyses, we established the null hypothesis (H_0) as the absence of a correlation between the two variables being compared, and the alternative hypothesis (H_1) as a correlation exists. We considered a p-value of less than 0.05 as the threshold for statistical significance. If the p-value fell below this threshold, it indicated sufficient evidence to reject the null hypothesis in favor of the alternative. We follow the general convention by Cohen [49] used in social sciences and consider a large correlation when the absolute value of the correlation coefficients are > 0.5, medium if they are around 0.3 and small if it is close to 0.1. However, correlations only characterize statistical pairwise co-occurrence and do not imply causation. By identifying correlations, we provide a critical understanding of related events that could guide further causal experimental research.

We complement the comparative analysis between absolute ranks with relative rank-difference analysis for social subgroups. Here, we determine the difference in ranking position for each activity between multiple subgroups or between a subgroup and the general population, providing a positive integer label l if the activity ranks l positions higher in the first analyzed subgroup than in the second and $-l$ if the activity ranks l positions lower. With these labels, we can create a new ranked list where the top activity is the one with the largest positive ranking changes, and the last is the one with the largest negative. An example can be seen in Fig. 4 with a comparison between DA for men and women (full ranking differences can be found on our website). Comparing these ranking changes between lists based on different variables (e.g., desired for automation vs. time), we can infer if the differences observed in one list correlate with the changes in another. For example, we can infer if activities with the most disparate ranking between men and women correspond to activities in which there is also a high degree of difference in the time-spent.

4 Results

We first compare the Desire for Robotic Automation (DA) to the time spent on activities, followed by comparing DA to well-being metrics. Table 1 summarizes correlations between the DA and the other dependent variables —Time spent (T), activity Happiness (H), Meaningfulness (M), Painfulness (P), Sadness (B), Stressfulness (S), and Tiredness (Z). Table 2 reports our results of analyzing the relative rank changes between DA and the other dependent variables for social subgroups —women (WN), men (MN), high (HI), middle (MI) and low income (LI)— when compared to the general population (GP).

4.1 Time as predictor for desire for robotic automation

Our analysis does not find a positive or negative correlation between the time spent on activities and the desire to automate them, neither in the general population nor in social subgroups (first row in Table 1). In all cases, the null hypothesis cannot be rejected ($p > 0.05$), indicating that T and DA may be completely independent. Fig. 3 depicts this lack of correlation graphically for the general population and the subset of *household activities*. The left-most plot shows the activities ordered by DA. The second-left plot depicts the time spent on the activities, using the same DA-based ordering: no clear pattern is visible. These results contradict our first hypothesis (**H1**).

Table 1: Rank Correlation of Desire for Automation

		GP	WN	MN	HI	MI	LI
DA-T	ρ	0.03 (.83)	−0.09 (.54)	0.03 (.83)	−0.05 (.72)	0.02 (.90)	−0.12 (.35)
	τ	0.03 (.73)	−0.06 (.50)	0.02 (.86)	−0.04 (.72)	0.02 (.87)	−0.09 (.38)
DA-H	ρ	**−0.36 (.01)**	−0.27 (.07)	**−0.36 (.01)**	−0.10 (.56)	**−0.29 (.04)**	−0.26 (.12)
	τ	**−0.26 (.01)**	**−0.20 (.05)**	**−0.24 (.02)**	−0.09 (.45)	**−0.19 (.04)**	−0.18 (.11)
DA-M	ρ	−0.21 (.15)	−0.21 (.17)	−0.25 (.08)	−0.23 (.20)	**−0.31 (.03)**	0.02 (.88)
	τ	−0.15 (.13)	−0.15 (.14)	−0.18 (.06)	0.13 (.27)	**−0.22 (.02)**	0.01 (.92)
DA-P	ρ	**0.30 (.04)**	−0.07 (.63)	0.27 (.07)	0.01 (.97)	0.22 (.13)	0.27 (.10)
	τ	**0.23 (.02)**	−0.05 (.60)	0.16 (.11)	0.03 (.82)	0.15 (.13)	0.20 (.09)
DA-B	ρ	0.08 (.57)	0.07 (.64)	−0.12 (.43)	0.02 (.93)	0.01 (.96)	0.09 (.58)
	τ	0.06 (.56)	0.04 (.71)	−0.06 (.52)	0.02 (.89)	−0.01 (.94)	0.08 (.51)
DA-S	ρ	0.22 (.13)	**0.32 (.03)**	0.02 (.88)	0.13 (.46)	0.11 (.44)	0.21 (.21)
	τ	0.16 (.11)	**0.22 (.04)**	0.01 (.92)	0.12 (.32)	0.07 (.47)	0.13 (.25)
DA-Z	ρ	0.07 (.62)	0.12 (.44)	−0.03 (.84)	0.11 (.52)	−0.01 (.96)	0.26 (.12)
	τ	0.04 (.67)	0.07 (.49)	−0.01 (.92)	0.10 (.41)	−0.01 (.90)	0.17 (.14)

Analyzing relative changes in rankings between DA and T for social subgroups (Table 2, first row), we observe a strong negative correlation between men and the general population, and between men and women. This suggests that when activities rank **higher** in DA for men (compared to the general population), they rank lower in T for men (compared to the general population). This trend is similar for ranked activities between men and women. This challenges hypothesis **H1b** for men: we assumed that a higher relative ranking in DA corresponds to a higher relative ranking in T.

4.2 Well-Being factors as predictors for desire for robotic automation

Happiness: Looking at the correlation between DA and H (second row, Table 1), we observe a medium negative correlation, indicating a low level of H experienced during an activity correlates to a high desire to automate it and vice versa. This is clear for the general population, men, mid-income subgroups and partially for women, where the probability of the null hypothesis is very low. Other subgroups show a similar trend, but the correlation is weak to medium and we cannot reject the null hypothesis. Based on these results, we consider a lack of Happiness a good indicator of the desire for robotic automation, especially for the general population, mid-income individuals and men (**H2.1**).

Analyzing relative changes in rankings between DA and H for social subgroups (Table 2, second row), the clearest pattern is between the high-income group and the general population, where we measure a strong positive correlation. This indicates that if an activity ranks higher in DA for the high-income subgroup than for the general population, it also ranks higher in H. For this group (and less marked, between middle-income and the general population and between high and middle-income), our results contradict our hypothesis (**H2.1b**) that activities ranking higher in DA will rank lower in Happiness.

Meaningfulness: Regarding the correlation between DA and M (third row, Table 1), the general population and most social subgroups tend to show a weak negative correlation, which supports our hypothesis (**H2.2**). We see an exception for the low-income subgroup, where the correlation is nearly zero. For all, except mid-income, the null hypothesis (no correlation) cannot be rejected.

When analyzing relative changes in rankings between DA and M for pairs of social groups (third row, Table 2), we do not observe clear patterns. An exception can be observed between the low-income group and the general population, where there is a medium positive correlation, indicating that if activities rank higher in DA for the low-income group (relative to the general population), they tend to also be ranked as more meaningful. Our hypothesis to explain changes in relative rankings as negatively correlated (**H2.2b**) to changes in M is not empirically supported.

Painfulness: For the general population and several social subgroups, we find a moderate positive correlation between DA and P (fourth row, Table 1), though the null hypothesis can only be rejected for the general population. Exceptions are women and high-income subgroups, where we measure a minimal correlation and the null hypothesis cannot be rejected. Thus, our hypothesis about a positive correlation explaining the relationship between DA and P (**H2.3**) holds for the general population.

When observing relative changes in DA rankings between pairs of social groups and their relationship to relative changes in the corresponding P rankings, we do not observe a clear pattern (fourth row, Table 2). Additionally, for most pairwise correlation analyses, the null hypothesis cannot be rejected. We do not find empirical support for our hypothesis about a positive correlation explaining changes in relative rankings (**H2.3b**).

Table 2: Relative Rank Correlation Analysis

		WN-GP	MN-GP	MN-WN	HI-GP	MI-GP	LI-GP
DA-T	ρ	−0.18 (.20)	**−0.43 (.00)**	**−0.42 (.00)**	−0.16 (.30)	0.11 (.44)	0.05 (.72)
	τ	−0.14 (.18)	**−0.31 (.00)**	**−0.29 (.00)**	−0.12 (.27)	0.08 (.47)	0.04 (.74)
DA-H	ρ	−0.02 (.91)	−0.10 (.51)	−0.13 (.42)	0.32 (.07)	0.23 (.13)	−0.02 (.92)
	τ	−0.01 (.93)	−0.08 (.48)	−0.08 (.49)	**0.25 (.05)**	−0.18 (.12)	−0.01 (.94)
DA-M	ρ	0.06 (.67)	0.01 (.93)	0.09 (.56)	0.04 (.84)	0.09 (.54)	**0.32 (.05)**
	τ	0.05 (0.66)	0.02 (.88)	0.07 (.52)	0.04 (.77)	0.07 (.53)	0.23 (.06)
DA-P	ρ	**0.31 (.04)**	0.00 (.99)	0.21 (.17)	−0.25 (.16)	0.14 (.36)	0.27 (.11)
	τ	0.21 (.06)	0.00 (.97)	0.13 (.23)	−0.15 (.22)	0.09 (.41)	0.20 (.10)
DA-B	ρ	0.15 (.33)	−0.27 (.07)	−0.22 (.14)	0.16 (.38)	−0.03 (.83)	0.08 (.63)
	τ	−0.09 (.41)	−0.19 (.09)	−0.16 (.15)	0.09 (.46)	−0.03 (.76)	0.06 (.60)
DA-S	ρ	−0.07 (.64)	0.11 (.47)	0.00 (.99)	−0.01 (.92)	0.00 (.98)	−0.11 (.51)
	τ	−0.04 (.75)	0.07 (.50)	0.00 (.99)	0.00 (.97)	−0.01 (.92)	−0.08 (.51)
DA-Z	ρ	0.19 (.22)	0.10 (.52)	0.22 (.16)	0.00 (.97)	−0.14 (.35)	0.07 (.70)
	τ	0.14 (.22)	0.08 (.48)	0.16 (.15)	0.00 (.97)	−0.10 (.38)	0.04 (.74)

Sadness: We find no clear correlation between DA and B either for the general population or for social subgroups (fifth row, Table 1), suggesting that Sadness is not a good predictor of automation desire (**H2.4**). The null hypothesis (no correlation) cannot be rejected.

Similarly, when observing the relative rank changes in DA for pairs of social groups and their correlation to rank changes in Sadness (row fifth, Table 2), we do not observe strong affinity, empirically invalidating our hypothesis of a positive relative correlation (**H2.4b**).

Stressfulness: We find a weak positive correlation between DA and S for the general population and most social subgroups (sixth row, Table 1). However, the null hypothesis can only be rejected for women, which also shows the strongest correlation. This supports our hypothesis that more stressful activities correlate with a higher desire for robotic automation, but only for women (**H2.5**).

When observing the relative rank changes in DA for pairs of social groups and their correlation with rank changes in stress (sixth row, Table 2), there are no strong correlations, contradicting our hypothesis of a positive correlation (**H2.5b**).

Tiredness: We find that Z has a weak, non-significant positive relationship with DA in the general population and in the social subgroups (seventh row, Table 1). For most subgroups, the null hypothesis probability is high, indicating that DA and Z being uncorrelated cannot be ruled out. Therefore, Tiredness is not a good predictor of the desire to automate, contrary to our initial hypothesis (**H2.6**).

When we observe the relative rank changes in DA for pairs of social groups and their correlation to rank changes in Z (seventh row, Table 2), we do not observe strong affinity with some correlations being weak and positive and others weak and negative, but with the null hypothesis being significant for all of them. Our hypothesis of a positive relative correlation was not supported (**H2.6b**).

5 Discussion

In the following, we analyze our research questions (RQ1-4) based on the results in the previous section. Our primary analysis is a statistical evaluation of activity ranks, complemented with illustrative or unexpected trends for particular activities.

RQ1: Does the average time-spent predict the desire for robotic automation? Our results contradict our first hypothesis, showing that time spent on an activity is **not** an indicator of the desire for robotic automation (**H1**). This contrasts prior work [50] that assumes people want to automate activities where they invest more time, motivating the search for other factors such as well-being. Remarkably, the activity with the lowest desire for robotic automation among the household subset is where people invest the most time: *food and drink preparation*. This raises questions about whether the numerous robotics efforts to automate cooking in industrial contexts would be successfully adopted in the home.

RQ2: Which feelings are the strongest predictors of the desire for robotic automation? Only the happiness and pain that activities bring to the general population are solid indicators for the desire to automate them (**H2.1** and **H2.3**). For many activities, the well-being scores for painfulness, sadness, stressfulness, and tiredness are close to zero, suggesting that these feelings were not evoked, rather than being low, which explains in part the lack of correlations. The meaningfulness and stress that the activities elicit are also weak to medium correlated —negatively (**H2.2**) and positively (**H2.5**), respectively—however we cannot reject the hypothesis that there is no correlation between them ($p > 0.05$). Interestingly, while *food and drink preparation* is associated with relatively high happiness and low desire for robotic automation among the household subset (last), *kitchen and food cleaning* ranks very low in happiness (18th) with a high desire for robotic automation: people seem to enjoy cooking but not cleaning afterwards, a task that robotics research could focus on.

RQ3: Do gender-based differences yield differences in the desire for robotic automation? While we observe significant differences in the time spent on activities and the desire to automate them based on gender, we only observe a clear pattern between the differences in time men spend on activities and their automation desires. Specifically, there is a strong negative correlation with respect to the general population and women (**H1b**), suggesting that *men exhibit a higher desire to automate activities they spend less time on when compared to women and the general population* and vice versa. We also observe that the correlation between the stress generated by the activities is more strongly (positively) correlated with the desire women have to automate them, which indicates that, for this social subgroup, stress is a good predictor: *women want to automate activities that stress them* (**H2.5**). Whereas *for men, low happiness serves as a more relevant predictor* (**H2.1**).

Remarkably, the largest differences between men and women in DA occur in activities associated with their stereotypical roles: men rank activities such as *laundry* and *sewing* much higher than women (largest negative rank difference in Fig. 4, right), corresponding to activities on which they spend much less time. In contrast, women rank activities such as *repairs*, *lawn/garden care* and *heating and cooling* higher than men in DA (largest positive rank difference in Fig. 4, right), corresponding to activities that rank lower than men in their spent time. A strong negative correlation between these differences indicates that the way men and women differ in how they prioritise automating activities corresponds inversely to the differences in the time they spend: *they want to automate the activities they spend less time on*, perhaps because they want *someone or something* to do these for them.

RQ4: Do socioeconomic differences yield differences in the desire for robotic automation? Overall, the trends observed in the general population tend to hold for the mid-income subgroup (fifth column of Table 1). There are some exceptions: while happiness and pain are good indicators of DA for the general population, happiness and meaningfulness serve as better indicators for middle-income individuals. When analyzing the way rank changes for different income subgroups, we do not observe clear patterns based on time spent (**H1b**) or well-being factors (**H2.Xb**) except in two surprising cases: 1) Differences in DA ranks between the high-income group and the general population are positively correlated with between-group differences in happiness rankings – indicating they want to automate activities that make them happier (Table 2, second row, fourth column). And 2) the differences in DA ranks between the low-income group and the general population correlate positively with the differences in meaningfulness ranks: activities that are more meaningful for individuals of the low-income group show a higher DA (Table 2, third row, sixth column).

Study Limitations: Firstly, the theoretical constructs used regarding the gendered division of labour assume heteronormative household structures. Also, our use of a binary gender classification reflects a broader issue in AI and HCI, where queer experiences are marginalised through classification [51]. Further, the demographic distribution of participants, particularly regarding race, limits our ability to conduct intersectional analyses that account for overlapping forms of oppression [52, 53], which is crucial given that women of colour disproportionately make up the housekeeping workforce and often shoulder a double burden when these activities become wage labor [54]. In particular, the number of high-income participants is significantly low. Since ATUS data is drawn from the U.S. population, our findings may not generalise to other cultural contexts. Finally, as an exploratory study engaging in open-science using three existing datasets, future work would benefit from collecting data from the same participants to directly relate time use, emotional experiences, and automation preferences.

6 Conclusion

We present an original study of the underlying reasons behind the desire for robotic automation, assessing how these tendencies change across subgroups with respect to time spent and feelings evoked. Our study integrated and analyzed data from three datasets, BEHAVIOR-1K, ATUS, and ATUS Well-Being Module. No correlation was found between automation desire and time spent, but happiness and pain were significant predictors. Significant trends indicated differences among genders and income levels. We open-source our data and an interactive analysis tool to enable reproducibility and support further robotics research.

Acknowledgements

This project was partially funded by Good Systems, an initiative of the University of Texas at Austin.

References

[1] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Michael Lingelbach, Jiankai Sun, et al. Behavior-1k: A benchmark for embodied ai with 1,000 everyday activities and realistic simulation. In *Conference on Robot Learning*, pages 80–93. PMLR, 2023.

[2] American Time Use Survey (ATUS). Year 2021. `https://www.bls.gov/tus/data/datafiles-2021.htm`, . Accessed: 2024-09-24.

[3] American Time Use Survey. The Well-Being Module (ATUS-WB). Year 2021. `https://www.bls.gov/tus/data/wbdatafiles-2021.htm`, . Accessed: 2024-09-24.

[4] Olga Miltsova. Robotic vacuum cleaner. iStock Photo by Getty Images, 2018. URL `https://www.istockphoto.com`. Standard license for personal, business, or commercial use.

[5] Arthur Rothstein. Knox county, tennessee. mrs. wiegel uses electric vacuum cleaner. Library of Congress, Prints & Photographs Division, FSA/OWI Collection, [LC-USW3-004045-D], 1942. URL `https://www.loc.gov`. No known rights restrictions, publication and other forms of distribution permitted.

[6] Silvia Federici. The reproduction of labour power in the global economy and the unfinished feminist revolution. *Workers and labour in a globalised capitalism: contemporary themes and theoretical issues*, 85, 2013.

[7] Steven M Gelber. Do-it-yourself: Constructing, repairing and maintaining domestic masculinity. *American quarterly*, 49(1):66–112, 1997.

[8] Megan Brenan. Women still handle main household tasks in us gallup. *Politics*, 2020.

[9] Sarah Jane Glynn. An unequal division of labor: How equitable workplace policies would benefit working mothers. *Center for American Progress*, 2018.

[10] Daniel Schneider and Orestes P Hastings. Income inequality and household labor. *Social Forces*, 96(2):481–506, 2017.

[11] Christine E Bose, Philip L Bereano, and Mary Malloy. Household technology and the social construction of housework. *Technology and Culture*, 25(1):53–82, 1984.

[12] Daniel Kahneman, Edward Diener, and Norbert Schwarz. *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation, 1999.

[13] Charles S Carver and Michael F Scheier. *On the self-regulation of behavior*. cambridge university press, 2001.

[14] James J Gross. Emotion regulation: Current status and future prospects. *Psychological inquiry*, 26(1):1–26, 2015.

[15] Hyunjeong Lee, Hyun Jin Kim, and Changsu Kim. Autonomous behavior design for robotic appliance. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 201–208, 2007.

[16] Maya Cakmak and Leila Takayama. Towards a comprehensive chore list for domestic robots. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 93–94. IEEE, 2013.

[17] Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. Virtualhome: Simulating household activities via programs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8494–8502, 2018.

[18] David Porfirio, Allison Sauppé, Maya Cakmak, Aws Albarghouthi, and Bilge Mutlu. Crowdsourcing task traces for service robotics. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 389–393, 2023.

[19] Nuno Otero, Aris Alissandrakis, Kerstin Dautenhahn, Chrystopher Nehaniv, Dag Sverre Syrdal, and Kheng Lee Koay. Human to robot demonstrations of routine home tasks: exploring the role of the robot's feedback. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 177–184, 2008.

[20] Guido Bugmann and Simon N Copleston. What can a personal robot do for you? In *Towards Autonomous Robotic Systems: 12th Annual Conference, TAROS 2011, Sheffield, UK, August 31–September 2, 2011. Proceedings 12*, pages 360–371. Springer, 2011.

[21] Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L Walters, Kheng Lee Koay, and Iain Werry. What is a robot companion-friend, assistant or butler? In *2005 IEEE/RSJ international conference on intelligent robots and systems*, pages 1192–1197. IEEE, 2005.

[22] Leila Takayama, Wendy Ju, and Clifford Nass. Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 25–32, 2008.

[23] JaYoung Sung, Henrik I Christensen, and Rebecca E Grinter. Sketching the future: Assessing user needs for domestic robots. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 153–158. IEEE, 2009.

[24] Eike Schneiders, Anne Marie Kanstrup, Jesper Kjeldskov, and Mikael B Skov. Domestic robots and the dream of automation: Understanding human interaction and intervention. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.

[25] Dahyun Kang, Sonya S Kwak, Hanbyeol Lee, and JongSuk Choi. First things first: A survey exploring key services and functions of a robot. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 278–280, 2020.

[26] Genevieve Bell and Joseph Kaye. Designing technology for domestic spaces: A kitchen manifesto. *Gastronomica*, 2:46–, 05 2002. doi: 10.1525/gfc.2002.2.2.46.

[27] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical hri. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 72–82, 2023.

[28] Suzanne M Bianchi, Liana C Sayer, Melissa A Milkie, and John P Robinson. Housework: Who did, does or will do it, and how much does it matter? *Social forces*, 91(1):55–63, 2012.

[29] Liana C Sayer. Gender, time and inequality: Trends in women's and men's paid work, unpaid work and free time. *Social forces*, 84(1):285–303, 2005.

[30] Michelle J Budig and Nancy Folbre. Activity, proximity, or responsibility? measuring parental childcare time. *Family time: The social organization of care*, 2:51–68, 2004.

[31] Michael Bittman, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. When does gender trump money? bargaining and time in household work. *American Journal of sociology*, 109(1):186–214, 2003.

[32] Jennifer L Hook. Women's housework: New tests of time and money. *Journal of Marriage and Family*, 79(1):179–198, 2017.

[33] Mark Aguiar and Erik Hurst. Measuring trends in leisure: The allocation of time over five decades. *The quarterly journal of economics*, 122(3):969–1006, 2007.

[34] Oriel Sullivan. What do we learn about gender by analyzing housework separately from child care? some considerations from time-use evidence. *Journal of Family Theory & Review*, 5(2): 72–84, 2013.

[35] Ekaterina Hertog, Setsuya Fukuda, Rikiya Matsukura, Nobuko Nagase, and Vili Lehdonvirta. The future of unpaid work: Estimating the effects of automation on time spent on housework and care work in japan and the uk. *Technological Forecasting and Social Change*, 191:122443, 2023.

[36] JaYoung Sung, Rebecca E Grinter, and Henrik I Christensen. Domestic robot ecology: An initial framework to unpack long-term acceptance of robots at home. *International Journal of Social Robotics*, 2:417–429, 2010.

[37] Maartje MA de Graaf, S Ben Allouch, and Jan AGM Van Dijk. What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*, pages 184–193. Springer, 2015.

[38] Hee Rin Lee and Selma Šabanović. Weiser's dream in the korean home: collaborative study of domestic roles, relationships, and ideal technologies. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, page 637–646. Association for Computing Machinery, 2013.

[39] Ja-Young Sung, Rebecca E Grinter, Henrik I Christensen, and Lan Guo. Housewives or technophiles? understanding domestic robot owners. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 129–136, 2008.

[40] Julia Fink, Valérie Bauwens, Frédéric Kaplan, and Pierre Dillenbourg. Living with a vacuum cleaning robot: A 6-month ethnographic study. *International Journal of Social Robotics*, 5: 389–408, 2013.

[41] Jodi Forlizzi and Carl DiSalvo. Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 258–265, 2006.

[42] Jennifer A Rode. The roles that make the domestic work. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 381–390, 2010.

[43] A Hochschild. *The second shift: Working families and the revolution at home*. Penguin Books, 2012.

[44] Catherine Bailey, Adrian Madden, Kerstin Alfes, Amanda Shantz, and Emma Soane. The mismanaged soul: Existential labor and the erosion of meaningful work. *Human Resource Management Review*, 27(3):416–430, 2017.

[45] Brent D Rosso, Kathryn H Dekas, and Amy Wrzesniewski. On the meaning of work: A theoretical integration and review. *Research in organizational behavior*, 30:91–127, 2010.

[46] Ruchira Ray, Leona Pang, Sanjana Srivastava, Li Fei-Fei, Samantha Shorey, and Roberto Martín-Martín. Why automate this? exploring correlations between desire for robotic automation, invested time and well-being. *arXiv preprint arXiv:2501.06348*, 2025.

[47] Historical poverty thresholds. URL `https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html`. Accessed: 2023-09-21.

[48] Pew Research Center. How the american middle class has changed in the past five decades, 2021. URL `https://www.pewresearch.org/social-trends/2021/07/01/how-the-american-middle-class-has-changed-in-the-past-five-decades/`.

[49] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[50] Céline Ray, Francesco Mondada, and Roland Siegwart. What do people expect from robots? In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3816–3821. IEEE, 2008.

[51] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, et al. Queer in ai: A case study in community-led participatory ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1882–1895, 2023.

[52] Sasha Costanza-Chock. *Design justice: Community-led practices to build the worlds we need*. The MIT Press, 2020.

[53] Eliza McCullough and Sarah Villeneuve. Participatory and inclusive demographic data guideline. *Partnership on AI*, 2024.

[54] Asha Banerjee, Katherine DeCourcy, Kyle K Moore, and Julia Wolfe. Domestic workers chartbook 2022: A comprehensive look at the demographics, wages, benefits, and poverty rates of the professionals who care for our family members and clean our homes.

[55] Morgan Klaus Scheuerman, Katta Spiel, Oliver L Haimson, Foad Hamidi, and Stacy M Branham. Hci guidelines for gender equity and inclusivity. 2020.

[56] Jessica J Cameron and Danu Anthony Stinson. Gender (mis) measurement: Guidelines for respecting gender diversity in psychological research. *Social and personality psychology compass*, 13(11):e12506, 2019.

[57] Greta R Bauer, Jessica Braimoh, Ayden I Scheim, and Christoffer Dharma. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PloS one*, 12(5):e0178043, 2017.

[58] Anna Brown, J Menasce Horowitz, Kim Parker, and Rachel Minkin. The experiences, challenges and hopes of transgender and nonbinary us adults. *Pew Research Center*, 2022.

[59] Banu Subramaniam, Rebecca Herzig, and Patrick Keilty. Queer data studies. 2024.

[60] Raj Korpan, Ruchira Ray, Andrea Sipos, Nathan Dennler, Max Parks, Maria E Cabrera, and Roberto Martín-Martín. Launching queer in robotics [women in engineering]. *IEEE Robotics & Automation Magazine*, 31(2):144–146, 2024.

[61] Katie Seaborn, Giulia Barbareschi, and Shruti Chandra. Not only weird but "uncanny"? a systematic review of diversity in human–robot interaction research. *International Journal of Social Robotics*, 15(11):1841–1870, 2023.