An Empathy-Based Sandbox Approach to Bridge Attitudes, Goals, Knowledge, and Behaviors in the Privacy Paradox

CHAORAN CHEN, University of Notre Dame, USA
WEIJUN LI, Zhejiang University, China
WENXIN SONG, The Chinese University of Hong Kong, Shenzhen, China
YANFANG YE, University of Notre Dame, USA
YAXING YAO, Virginia Tech, USA
TOBY JIA-JUN LI, University of Notre Dame, USA

The "privacy paradox" describes the discrepancy between users' privacy attitudes and their actual behaviors. Mitigating this discrepancy requires solutions that account for both system opaqueness and users' hesitations in testing different privacy settings due to fears of unintended data exposure. We introduce an empathy-based approach that allows users to experience how privacy behaviors may alter system outcomes in a risk-free sandbox environment from the perspective of artificially generated personas. To generate realistic personas, we introduce a novel pipeline that augments the outputs of large language models using few-shot learning, contextualization, and chain of thoughts. Our empirical studies demonstrated the adequate quality of generated personas and highlighted the changes in privacy-related applications (e.g., online advertising) caused by different personas. Furthermore, users demonstrated cognitive and emotional empathy towards the personas when interacting with our sandbox. We offered design implications for downstream applications in improving user privacy literacy and promoting behavior changes.

CCS Concepts: • **Security and privacy** → **Human and societal aspects of security and privacy**; • **Human-centered computing** → *Human computer interaction (HCI).*

Additional Key Words and Phrases: privacy paradox, privacy intervention, decision making, empathy

1 INTRODUCTION

Privacy paradox [59, 75] is a common phenomenon that refers to the discrepancy between the attitude of users and their actual behaviors in managing their privacy. This inconsistency has been observed in various domains, including social networking service [49, 76, 104], online shopping [8, 10, 17], mobile app [13, 79, 85] and Internet-of-Things [101]. Prior work in behavioral economics and decision research found several cognitive and behavioral biases that lead to the privacy paradox, such as hyperbolic discounting [46], the immediate gratification [1], and the illusion of control [20, 21].

However, bridging the privacy paradox can be challenging for two reasons. From a system perspective, the inherent *opaqueness* in the system barriers users from making informed decisions about protecting their privacy. The asymmetric information [3] provided by the system makes it challenging for users to understand what data is collected and how other parties use it [64, 67, 105]. Consequently, users are unable to make informed decisions to safeguard their personal data while maintaining the desired level of usability and system utility, such as whether to opt out of certain data collection practices, configure the frequency and granularity of data sharing, or the adoption of privacy-enhancing tools. From the user perspective, the *fear of exposing personal data* [63, 93, 108] while navigating an opaque system can further discourage users from experimenting with different possible privacy configurations to link their available

Authors' addresses: Chaoran Chen, cchen25@nd.edu, University of Notre Dame, Notre Dame, Indiana, USA; Weijun Li, Zhejiang University, Hangzhou, China; Wenxin Song, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China; Yanfang Ye, University of Notre Dame, Notre Dame, Indiana, USA; Yaxing Yao, Virginia Tech, Blacksburg, Virginia, USA; Toby Jia-Jun Li, University of Notre Dame, Notre Dame, Indiana, USA.

options of privacy choices to their consequences, thereby reinforcing the system's opaqueness. Once users share their private data, they will no longer have control over how the other party utilizes the information [78]. Besides, another barrier that prevents users from meeting their privacy goals is their *limited experience and lack of privacy literacy*. Lay users are prone to perceive fewer privacy threats compared with technicians [58]. Thus, even if users sometimes know their privacy goals, they still trade off their privacy for convenience, as they believe that their data is well-protected by the system.

To support users' privacy decision-making, two approaches have been widely applied: privacy education and nudging [102]. Privacy education endeavors to cultivate privacy awareness and literacy, thereby equipping users with the knowledge to make well-informed privacy choices. However, a notable challenge with these methods lies in the extended duration required for shifts in privacy attitudes. Users often encounter difficulties in adhering to expert privacy recommendations and translating acquired knowledge into specific online contexts. An alternative approach to facilitating privacy decisions is through privacy nudging[2]. Nudges encompass subtle yet influential prompts that steer individuals toward certain behaviors. Although nudges can facilitate the adoption of specific behaviors, their effects tend to be transitory, as intermittent adjustments in individuals' privacy practices may not necessarily extend to their overall privacy literacy.

To address the aforementioned challenges and limitations, we present an *empathy-based* method that allows users to experience and observe the correlation between their privacy data and the system outcomes in a real-time and risk-free environment. In this approach, we use *personas* that come with synthesized personal data based on real-world privacy incidents. Each persona represents a fictional user [19] with a distinctive biography, demographic information, and a large set of synthesized personal data. For example, here is an exemplary biography of a "tangible" persona:

- Alice is a 40-year-old white woman living in New York. She is an administrative assistant, and her annual income is around ninety thousand USD. She lives with her husband and two teenage children. She is an avid user of social media platforms such as Facebook and Instagram, where she often shares posts, photos, and videos of her life. She also often purchases clothing items and books on Amazon.

Unlike personas often used in the user experience design process, personas used in our context should also include plausible *realistic* longitudinal personal data such as web browsing history, social media logs, location records, and weekly schedules. The intricate realness of these personas is facilitated by the use of Large Language Models (LLMs), which can generate a diverse range of highly detailed and modifiable personas. Our design draws upon the principles of empathy-based design. Recognized for its essential role in user experience and persuasive design, the empathy-based design employs narrative and role-play techniques to establish deeper and more meaningful connections with users [24, 32, 103].

Through the sandbox, users can interact with different online services, as usual, using the identities of their selected personas. The sandbox will be loaded with personal data from the persona instead of the user, so whenever an online service queries personal data, the synthetic personal data associated with the persona will be provided. As far as service providers are concerned, the data appear real, causing them to offer personalized content and services as though interacting with the user the persona represents. This gives users a risk-free platform to investigate privacy settings and actions, perceive the resulting user experience, notice the tangible consequences of their privacy choices, and experience emotional results, positive or negative, in a convincingly interactive environment without exposing their actual personal data.

We validate the proposed approach through a prototype (i.e., Privacy Sandbox) and a study involving 15 participants. The results validated the technical feasibility of our approach to generate artificial personas with realistic synthesized personal data. Our findings imply that users can indeed establish empathy with personas when using the Privacy Sandbox and identify links between the persona's privacy attributes with the observed system outcomes. The results of the study also offer design implications for using the proposed approach to empower users to acquire privacy knowledge and promote behavior change.

This paper makes the following contributions:

- Introduces an empathy-based approach that allows users to experience the links between privacy behaviors and system outcomes in a risk-free sandbox environment using artificially generated personas.
- Validates the viability of the proposed approach through a prototype implementation and empirical studies. The study results confirmed the users' cognitive and emotional empathy toward the generated personas when interacting with the sandbox in the context of target advertisements.
- Discusses the design implications of adopting this empathy-based privacy persona approach to empower users to acquire privacy knowledge that leads to behavior change.

2 RELATED WORK

2.1 Empirical studies on privacy paradox

The existence of the privacy paradox is a long-lasting debate in privacy studies. Some researchers regard it as a "myth" [92] because the behavior in the privacy paradox studies pertains to the decision-making about risks in very specific contexts. In contrast, their self-reported privacy concerns are much more general. However, many empirical studies [4, 29, 59] have consistently demonstrated the existence of the privacy paradox across various distinct scenarios.

Most empirical studies measure the gap between users' privacy attitudes and behaviors, commonly using surveys. For example, Madejski et al. [65] used surveys to gauge privacy attitudes, previous privacy settings, and self-reported sharing intentions on Facebook, identifying potential sharing violations by comparing intentions with settings. Colnago et al. [30] also employed within-subjects surveys, revealing mismatches between attitudes/preferences and behaviors.

Although surveys are able to explore privacy attitudes, they are not reliable when examining irregular or infrequent privacy behavior [59, 94]. Consequently, many studies combine surveys with experiments for more reliable behavior data collection. For instance, Norberg et al. [75] assessed willingness to disclose information in surveys and later conducted a field study to compare willingness with actual disclosure, finding significant differences. Barth et al. [13] measured privacy concerns through surveys and compared results with participants' actual behavior, represented by the number of intrusive apps downloaded.

While multiple studies have measured the privacy attitude-behavior gap, only a few have proposed ways to address it. Previous research has examined risk awareness [36], the privacy calculus [77], and digital nudges [56] as potential solutions. Sutanto et al. [96] designed a personalized privacy-safe application that retains users ' information locally on their smartphones while still providing them with personalized products. Mattson et al. [66] suggested changing negative attitudes in different functional areas to reduce the intention-behavior gap.

Unlike previous methods relying on surveys and/or experiments for on-the-spot decisions, our approach offers a risk-free environment for users to experience online services by using different generated personas' data. This allows users to reflect on system outcomes and make more informed decisions, potentially mitigating the privacy paradox gap resulting from ad-hoc decisions.

2.2 Approaches to prompt privacy behavior change

Previous theories on privacy behavior [5, 11, 62] emphasize the role of attitudes in shaping behavior. Based on them, we distinguish between two approaches in promoting privacy behavior change: the *top-down approach*, which primarily seeks to foster privacy attitudes and literacy to indirectly influence behavior, and the *bottom-up approach*, which directly influences behavior through techniques like nudging. We aim to comprehensively assess their effectiveness and improve their impact on privacy behavior.

Top-down approaches (e.g., privacy education) aim to improve privacy behavior by fostering privacy attitudes [6, 39] and enhancing privacy literacy [43]. For instance, Desimpelaere et al. [35] observed that privacy literacy training improves children's understanding and promotes privacy-protective behavior. Sideri et al. [88] found that university-based education enhanced students' digital knowledge and privacy awareness on social networking sites. Innovative methods like Franco et al. [44] employed technology-enhanced pedagogical scenarios to involve students in active learning by using their own social media traces. Despite increased awareness and knowledge, top-down approaches may not consistently translate into behavior change. Users often struggle to apply experts' privacy advice [33] due to its vagueness [83] and lack of alignment with their specific needs and contexts [107].

Bottom-up approaches (e.g., digital nudging) alter user privacy behavior by guiding their choices in digital environments using techniques such as visual cues [86], information presentation [27], default settings [9], and incentives [7]. These nudges can reduce data disclosure and influence privacy choices in a short period [27, 40]. While privacy nudges can alter privacy behavior quickly, they often lead to temporary effects as they may not necessarily improve users' long-term privacy literacy or decision alignment with their attitudes.

Our approach combines both top-down and bottom-up approaches. It offers a systematic risk-free platform for users to learn about privacy in a structured way (top-down) while providing experiential learning [45] through real-world consequences based on user interactions and privacy choices (bottom-up). This dual approach aims to bridge the gap between privacy knowledge and behavior effectively.

2.3 The use of empathy in user experience design and persuasive design

Empathy, often defined as the capacity for an affective response aligned with someone else's situation rather than one's own [14], encompasses both affective and cognitive components. *Affective empathy* involves an immediate emotional response to others, while *cognitive empathy* pertains to understanding others' feelings [32, 37, 42, 60]. It is a powerful instrument to connect people with others and has been applied in various domains like user experience (UX) design [32] and persuasive design.

In UX field, empathic design [60] aims to enable designers to 'step into the user's shoes' and 'walk the user's walk', thereby crafting products that align with user needs. A solid foundation for comprehending empathy in design research is established through the exploration of philosophy, psychology, and neuroscience literature. Surma-Aho et al. [95] offered a comprehensive review of empathy's role in design. A pertinent framework, proposed by Hess and Fila [54], defines empathy along two axes: affective experiences vs. cognitive processes and self-oriented vs. other-oriented perspectives, yielding four dimensions:

(1) *perspective-taking*, where designers imagine users' thoughts and feelings (cognitive, other-oriented);
(2) *empathic concern*, as designers display sincere care for users (affective, other-oriented);
(3) *emotional congruence*, with designers sharing users' emotional states (affective, self-oriented);
(4) *projection*, when designers experience unease due to users' challenges (cognitive, self-oriented).

Building on it, various approaches, such as narrative and role-play techniques [103], have been developed to foster deeper connections with users and their experiences in empathic design. These methods involve creating scenarios and personas to envision potential design innovations [28] or simulating user experiences through role-play [24].

Our work draws inspiration from these empathic understanding frameworks and design methods, aiming to investigate whether users can develop empathy toward the generated personas and whether this influences their acquisition of privacy knowledge.

In persuasive design, stimulating empathy is a crucial technique [26]. Previous studies have harnessed the malleable nature of empathy to promote prosocial behavior [31, 89, 98]. For example, S.H. Taylor et al. [98] found that embedding empathy nudges in social media posts can encourage bystander intervention for cyberbullying victims. Additionally, many researchers have found that designing with empathy can encourage the natural empathetic behavior of people who have existing social ties or shared interests [80, 98] A typical example is that VR can enhance cognitive empathy by emphasizing user similarities [87].

Our motivation is aligned with the concept of bystander empathy [98], aiming to modify user behavior by eliciting their empathy towards generated personas. To achieve this goal, we draw upon empathy-inducing techniques from persuasive design, such as providing detailed and specific information [31, 90], immersive role-playing and perspective-taking [16, 87], and considering connections between users and personas [87].

While various measurement scales exist to quantify empathy in psychology [12, 34, 41, 55], our focus is on the relationship between users and personas rather than personal characteristics. We derive inspiration from approaches used to measure designers' empathy during the design process, including indicators like empathic expressions, personal experiences, respectful questioning, and discussing user facts [99]. Our empathy measurement approach combines self-report methods and integrates established theoretical frameworks of empathy.

3 AN EMPATHY-BASED PRIVACY PERSONA APPROACH

3.1 Overview

In an effort to bridge the gap between users' attitudes and their behaviors in managing their privacy, we introduce a new empathy-based sandbox approach. This approach uses artificially generated user personas with realistic synthesized personal data, enabling users to (1) load synthetic personal data into browsers; (2) interact with websites and applications "under the disguise" of an artificial persona; (3) experiment with various privacy configurations and behaviors in a risk-free environment; and (4) experience the corresponding outcomes (both positive and negative).

As illustrated in Fig. 1, the key goals of our empathy mechanisms are two-fold.

(1) **Emotional Resonance**: When users encounter privacy incidents when acting as personas, we believe that users can feel the emotion that the persona would have felt (e.g., frustration or anger from privacy violations; joy from apt personalized recommendations). Users will also realize how they may feel when they encounter a similar incident.

(2) **Knowledge Acquisition**: When users experience an outcome of their privacy behaviors (e.g., seeing a particular personalized ad or being influenced by an algorithmic decision) when acting as personas, we believe that users will be able to identify patterns and acquire generalizable knowledge, which fosters a more intuitive understanding of likely outcomes from their future actions.

As discussed in Section 2, it is feasible to stimulate users' empathy toward privacy personas by employing empathy-inducing techniques from user experience design [103] and persuasive design [98]. Simultaneously, users can learn

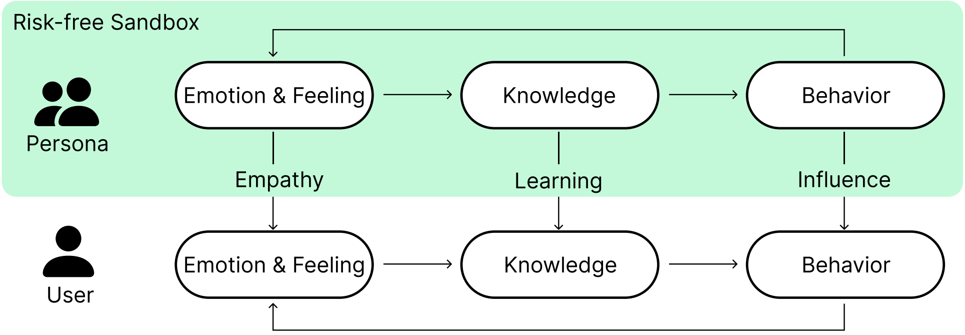


Fig. 1. An empathy-based approach where users interact with online services with different personas in a risk-free sandbox without leaking their real personal data. Users can observe and experience the causal effect between their privacy configurations/behaviors and system outcomes, acquire privacy knowledge, and translate the knowledge into actual behavior.

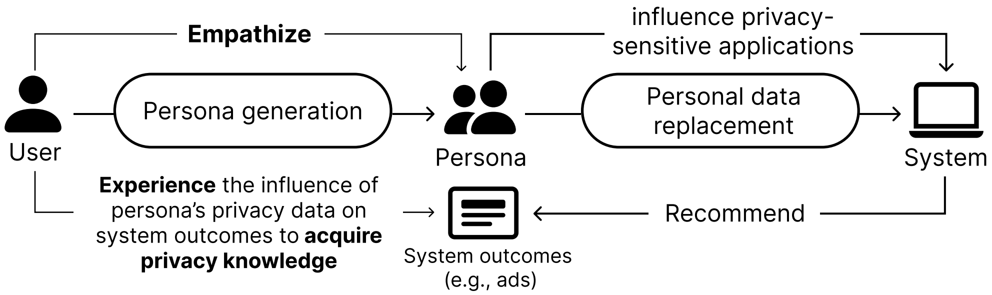

Empathize
influence privacy-sensitive applications
User
Persona generation
Persona
Personal data replacement
System
Experience the influence of persona's privacy data on system outcomes to acquire privacy knowledge
System outcomes (e.g., ads)
Recommend

Fig. 2. An empathy-based approach where users interact with online services by using the identity of different personas in a risk-free sandbox without leaking their real personal data. Users can cognitively and emotionally empathize with personas, observe and experience the causal effect between the privacy data and system outcomes (e.g., target ads), and acquire privacy knowledge.

about privacy in a structured and interactive manner by experiencing the influence of privacy personas' information on system outcomes. Such experiential learning not only provides real-time feedback, akin to nudging techniques [52] but also facilitates users to acquire privacy knowledge [43]. Therefore, combining emotional resonance with privacy knowledge acquisition can result in more *motivated* and *informed* users. Rooted in current frameworks and past research, we hypothesize that fostering both emotional resonance and knowledge acquisition can promote privacy literacy, subsequently leading to changes in user privacy behaviors that align with their preferences.

Our approach has two core phases: persona generation and personal data replacement. Persona generation involves constructing personas with sufficient detail. Once completed, in personal data replacement, users will use personas to explore various privacy settings and online services from the perspective of a particular persona. An overview of this methodology can be seen in Fig. 2. In the following sections, we explain the two steps in detail, followed by introducing a prototype that integrates both steps—the Privacy Sandbox.

3.2 Persona generation

The goal of the persona generation stage is to create artificial personas that contain realistic synthetic personal data. By doing this, external applications and web services will read the synthetic data of the personas instead of the real data of actual users when they use the interactive sandbox. Consequently, these applications and services will tailor content based on the generated persona, letting users see the results of different privacy behaviors without risking their actual data.

To effectively influence system outcomes and invoke user empathy, we chose specific data attributes to include in our personas.

3.2.1 *Selected data attributes.*

- *Personally identifiable information (PII):* first name, last name, profile picture, and date of birth.

Our rationale for choosing these attributes is as follows. Although recommendation systems are often based on anonymized data [84] and do not rely on profile pictures, names and profile pictures are still fundamental in personas [48]. They make personas real and relatable, serving as vital stimuli of user empathy [38]. Furthermore, birth dates not only help establish the persona's age, making them more recognizable to a certain age group, but also are pivotal for personalized content [68]. However, due to ethical concerns, we omitted sensitive PII like phone numbers and Social Security Numbers.

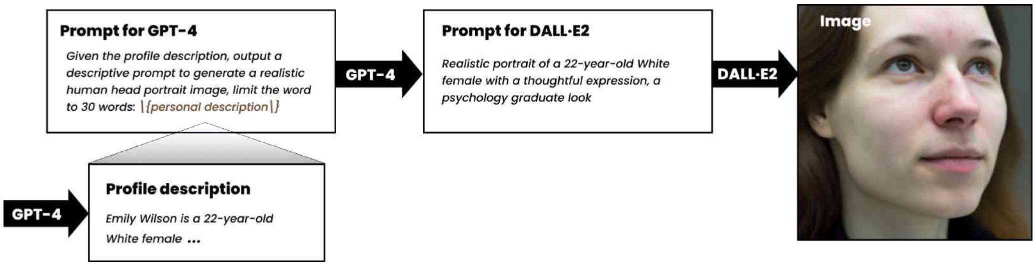
- *Demographic information:* age, gender, race and ethnicity, languages, education, income, occupation, home address, marital status, and parental status.

Demographic details, hobbies, and online interests play a crucial role in creating realistic personas. These attributes allow users to quickly connect with and relate to the personas through shared characteristics [68]. Such connections foster user engagement and empathy, making interactions with personas more meaningful and relatable. Furthermore, online recommendation systems often utilize demographic data and personal preferences to tailor their offerings [70]. This kind of personal data guides how online platforms categorize users and, subsequently, the type of content they receive. When users observe how demographic information and personal preferences impact the services or user experiences, they are more likely to disentangle the system's opaqueness, understand how the system might use their privacy information, and consequently enhance their privacy awareness and literacy. Additionally, these data contribute significantly to the generation of longitudinal personal data. They influence a user's weekly schedule by reflecting their lifestyle choices and priorities. These data also shape one's browsing history and social media as individuals tend to browse and share content related to their hobbies and demographic identities.

- *Additional personal information:* devices in use, browser in use, hobbies, and online interests.

We incorporated details about devices and browsers into our personas, given their influence on online service personalization. Notably, certain studies, such as the one by Nikiforakis et al. [74], found that advertisers can use browser and device fingerprinting to tailor the ads for users. Hannak et al. [51] observed that e-commerce platforms might offer different prices based on the user's device or browser. Recognizing such distinctions can help users understand the link between their device/browser information and the system outcomes, which consequently enhances their privacy knowledge.

- *Longitudinal personal data:* weekly schedule with location logs, browsing history, and social media posts.



Given the profile description, output a descriptive prompt to generate a realistic human head portrait image, limit the word to 30 words: \{personal description\}

Realistic portrait of a 22-year-old White female with a thoughtful expression, a psychology graduate look

Emily Wilson is a 22-year-old White female ...

Fig. 3. The generation pipeline of profile portrait images.

Differing from traditional UX personas, we incorporated longitudinal personal data, like weekly schedules, location logs, browsing histories, and social media activities. From the user's perspective, this richer dataset paints a more comprehensive picture of the persona's life, fostering deeper user empathy. Previous work [48] has also underscored the significance of longitudinal personal data in understanding personas. Rijn et al.'s study [99] emphasizes the value of in-depth behavioral data in understanding personas. Thus, by observing these personas' daily activities and interests, users can better empathize with them, seeing them as dynamic individuals with changing preferences.

From the system's perspective, this temporal data is pivotal for online services to make contextual recommendations. Several studies have illustrated the use of social media [25] and browsing histories [15, 82, 97] in predicting user preferences. Consequently, this data not only amplifies user empathy but also impacts the tailored content they encounter. This insight helps users grasp the connection between their data and the content they receive, boosting their privacy awareness and possibly guiding their future privacy behaviors.

3.2.2 Data generation methodology. To create comprehensive persona data, we introduce a novel pipeline that augments the outputs of large language models using few-shot learning, contextualization, and chain-of-thoughts techniques. For readers' reference, we have included detailed prompts and examples of few-shot learning in Appendix A.

Persona description: The foundational step in our process is generating a personal description, which informs subsequent data generation to ensure alignment. We utilize a template prompt coupled with few-shot learning [23], to guide GPT-4 in producing personally identifiable and demographic information. Users can customize the generation of their desired persona by providing other guidance as input.

Privacy attributes: We use GPT-4 and few-shot learning to parse the generated persona description and obtain attributes for each PII and demographic information to allow further modifications.

Profile portrait image: To make the generated persona feel more tangible and authentic to users, we employ a "chain-of-thought" approach [100] to create prompts for the generation of profile portrait images. As shown in Fig. 3, we start by entering the personal description to generate a prompt for the OpenAI DALL·E 2 image generation API [1]. After obtaining the prompt, we then invoke the image generation API to synthesize a personal portrait image.

Device and browser: Since device and browser information is typically contained within the browser's user agent, and user agent information does not appear directly in the personal description, we have separately created a prompt for GPT-4 to predict user agent, device, and browser based on the persona's personal description.

Weekly schedule with location records: A persona's weekly schedule provides insights into daily routines, further informing the creation of browsing histories and social media posts. By employing few-shot learning and

[1] https://platform.openai.com/docs/guides/images/image-generation-beta

contextualization, the persona's description is embedded within the prompt, ensuring schedule consistency. For geographical context, we have incorporated sample addresses into the few-shot learning examples, ensuring that generated events include reasonable location information.

Browsing history: To generate consistent browsing history that aligns with the persona's personal description and weekly schedule with location records, we include them as the context in the prompt for generating browsing history for a specific time period. During the generation process, we also utilize few-shot learning to provide sample references for the browsing history records.

Social media posts: To ensure realistic and consistent social media posts, we use the persona's description and weekly schedule as contextual anchors in the prompt. Typically, social media posts might contain visuals, so we randomly add 0-2 images per post. If an image is integrated, we employ the "chain-of-thought" technique similar to how we generate profile pictures: the post content serves as input, generating a prompt for image synthesis. This prompt is then fed to the OpenAI DALL·E image generation API. This process enriches the realism of the persona, aligning visual content with the textual post. An illustrative prompt for social media posts is presented below.

Prompt for generating social media posts

Provide ideas for this person to write posts (limit the word to 140 words) based on the profile and location history: {profile} {location history}
Return a list of lists: <few-shot example posts>
Output the posts in the following JSON format in plain text: { "time": <time in string format>, "address": <address where this person shares the life>, "content": <content>, "latitude": <fake latitude>, "longitude": <fake longitude>, "timezone": <time zone>, "locale": <locale> }

Few-shot learning example for the generation of social media posts

"posts": "[["2023-06-01 08:31:10", "Starting my day with a delicious cup of coffee at my favorite coffee shop. Ready to conquer the world! #CoffeeLover #MorningMotivation", "Coffee Shop - 123 Main Street, Brooklyn, New York 11207"], ["2023-06-01 18:00:34", "Just got back from the grocery store. Stocked up on essentials for the week. #GroceryHaul #MealPrep", "Grocery Store - 456 Broadway Avenue, Brooklyn, New York 11207"]]"

Prompt for generating a prompt for generating an image associated with the post

Given the post {content}, output a descriptive prompt to generate a realistic life image, limit the word to 30 words:

3.2.3 Implementation details. To generate synthetic data, we used a Python script that interacted with the GPT-4 public API. We specified a maximum continuation length of 4, 500 tokens. Our approach to achieving few-shot learning involved utilizing the "FewShotPromptTemplate" available in the open-source Python library called "langchain[2]". Furthermore, we configured the GPT-4 model with a temperature parameter of 0.9. The resulting images, generated using OpenAI DALL·E, were set to a size of 256 × 256 pixels.

[2] https://pypi.org/project/langchain/

3.3 Personal data replacement

In the second stage, users can use the identities of the generated personas to interact with various online services. The sandbox replaces users' demographic data within the Google account, real-time location, IP address, and web browsing history to match the persona's attributes. When an online service requests this personal information, the system offers the synthetic data of the persona. For the service provider, this data seems genuine, allowing them to provide personalized content as though they were interacting with a real user. This approach offers users a risk-free platform to cognitively understand the tangible consequences of their privacy choices and emotionally empathize with the persona in a convincingly interactive environment without exposing their actual personal data. The subsequent sections detail the process of personal data replacement.

Google account: Since we choose online ads to represent system outcomes, data replacement for Google accounts primarily pertains to information within the Google Ad Center. Google Ad Center's control portal[3] allows users to customize the information provided to Google Ads, encompassing details of age, gender, language, relationship status, household income, education, industry, and homeownership. To substitute the profile data in the Google Ad Center with the privacy data of the persona, we create a Google account dedicated to the application. We use three open-source node.js libraries ("Puppeteer[4]", "Puppeteer-extra", and "Puppeteer-extra-plugin-stealth[5]") to automate the replacement of profile data. The replacement process consists of three steps: (1) After entering the personal profile page in Google Ad Center, we traverse the "aria-labels" of all elements of the page to identify the attributes that need to be replaced. (2) We extract the persona attributes from the database and process the data. (3) Then, we replace the values of target attributes with the persona's corresponding information.

Geographical location: Personalized online advertisements are often tailored based on the user's geographical location. Chrome browser supports location override. Replacement of geographical location involves two steps. First, based on the persona's home address, we use the open-source mapping application OpenStreetMap's geocoding API[6] to obtain the latitude and longitude of the generated persona's current address (based on their generated schedule). Then, we use the "setGeolocation" method from Puppeteer to modify the geographical location of the webpages based on the obtained latitude and longitude coordinates.

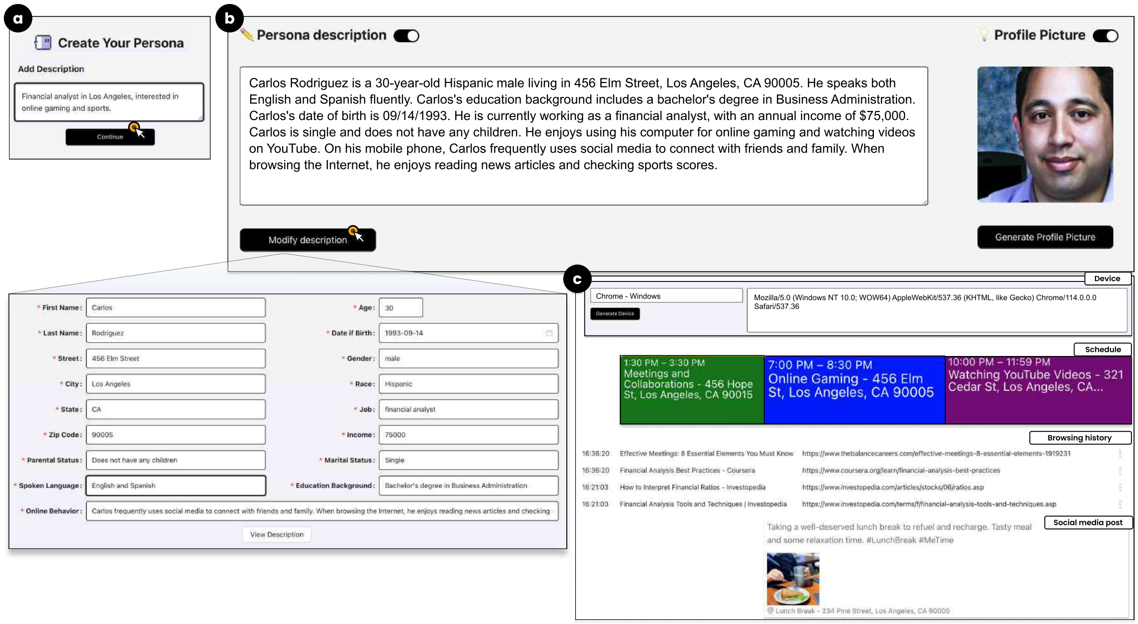
IP address: We use NordVPN's API[7] to modify the user's IP address. Based on the latitude and longitude of the persona's current location we obtained from OpenStreetMap, we calculate the nearest NordVPN server station to that location and select the server with the lowest load for connection. Once this connection is made, any online service that inquires about the IP address will receive the server's IP address instead of the user's original one.

User agent: To adjust the user agent, we employ the "setUserAgent" function in Puppeteer. This replaces the current page's user agent with the device and browser details associated with the generated persona.

Browsing history: Chrome keeps its browsing history on the local computer using the SQLite database. Before launching the browser, we utilize a JavaScript script to overwrite the corresponding database file. Specifically, we substitute both the URL table, which logs visited links, and the visit table, which notes browsing timestamps, with the browsing history of the generated persona.

3.4 A prototype for study: Privacy Sandbox

[3] https://myadcenter.google.com/controls
[4] https://pptr.dev/
[5] https://github.com/berstend/puppeteer-extra
[6] https://nominatim.openstreetmap.org/ui/search.html
[7] https://nord.readthedocs.io/en/latest/reference/api.html



Financial analyst in Los Angeles, interested in online gaming and sports.

Carlos Rodriguez is a 30-year-old Hispanic male living in 456 Elm Street, Los Angeles, CA 90005. He speaks both English and Spanish fluently. Carlos's education background includes a bachelor's degree in Business Administration. Carlos's date of birth is 09/14/1993. He is currently working as a financial analyst, with an annual income of $75,000. Carlos is single and does not have any children. He enjoys using his computer for online gaming and watching videos on YouTube. On his mobile phone, Carlos frequently uses social media to connect with friends and family. When browsing the Internet, he enjoys reading news articles and checking sports scores.

* First Name:	Carlos	* Age:	30
* Last Name:	Rodriguez	* Date if Birth:	1993-09-14
* Street:	456 Elm Street	* Gender:	male
* City:	Los Angeles	* Race:	Hispanic
* State:	CA	* Job:	financial analyst
* Zip Code:	90005	* Income:	75000
* Parental Status:	Does not have any children	* Marital Status:	Single
* Spoken Language:	English and Spanish	* Education Background:	Bachelor's degree in Business Administration
* Online Behavior:	Carlos frequently uses social media to connect with friends and family. When browsing the Internet, he enjoys reading news articles and checking		

Chrome - Windows

Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/114.0.0.0 Safari/537.36

1:30 PM – 3:30 PM Meetings and Collaborations - 456 Hope St, Los Angeles, CA 90015	7:00 PM – 8:30 PM Online Gaming - 456 Elm St, Los Angeles, CA 90005	10:00 PM – 11:59 PM Watching YouTube Videos - 321 Cedar St, Los Angeles, CA...

16:36:20	Effective Meetings: 8 Essential Elements You Must Know	https://www.thebalancecareers.com/effective-meetings-8-essential-elements-1919231
16:36:20	Financial Analysis Best Practices - Coursera	https://www.coursera.org/learn/financial-analysis-best-practices
16:21:03	How to Interpret Financial Ratios - Investopedia	https://www.investopedia.com/articles/stocks/06/ratios.asp
16:21:03	Financial Analysis Tools and Techniques | Investopedia	https://www.investopedia.com/terms/f/financial-analysis-tools-and-techniques.asp

Taking a well-deserved lunch break to refuel and recharge. Tasty meal and some relaxation time. #LunchBreak #MeTime

Lunch Break - 234 Pine Street, Los Angeles, CA 90005

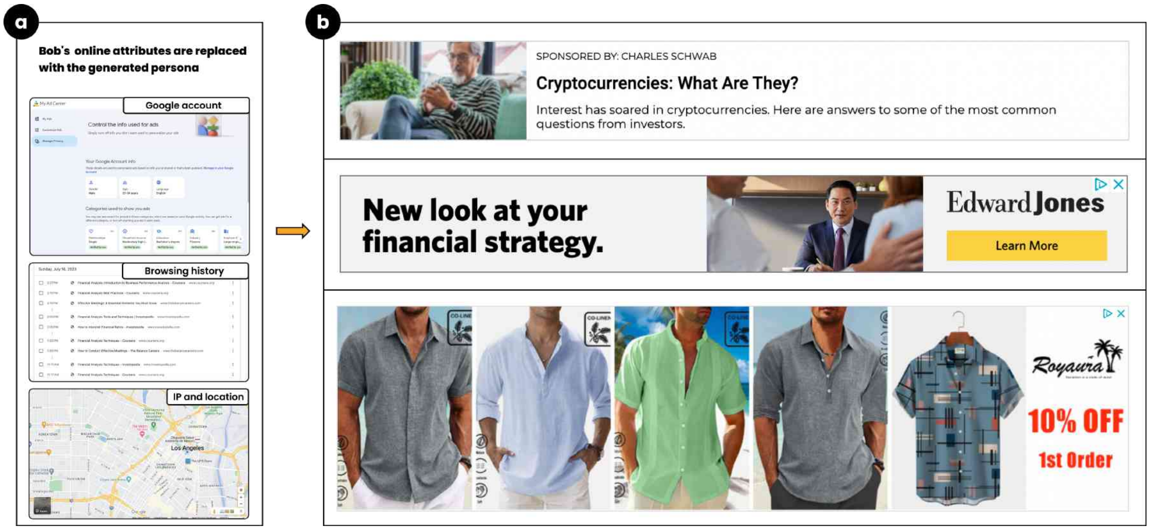
Fig. 4. Privacy Sandbox User Journey. (a) Providing guidance for Persona's Profile Generation: The User's initial input acts as a seed for persona creation, exemplified by Bob's specific professional and personal interests. (b) Initial Persona Profile Generation and Customization: Creation of a preliminary persona "Carlos Rodriguez", which users can review and modify. (c) Generating additional privacy data aligned with the profile: Extension of the persona's attributes, ensuring alignment with the initial profile.

3.4.1 Privacy sandbox in action. We demonstrate the use of the Privacy Sandbox through an example usage scenario. In this scenario, a user creates a persona to navigate online services, showcasing the core features of Privacy Sandbox.

Consider financial analyst Bob who wants to understand how private data impacts online ads. Using the Privacy Sandbox, he can generate a persona and act as the persona to browse websites that contain ads.

(1) *Providing Guidance for Persona Profile Generation:* Bob chooses the "create a new persona" button and inputs his guidance to generate a persona that is similar to his profile. He enters "Financial analyst in Los Angeles, interested in online gaming and sports." The "guidance" in this context acts as a seed or initial information. Users can provide as little or as much information as they feel comfortable with, ensuring flexibility while guarding their own private information. This information is not restricted to job titles or locations but could include hobbies, interests, or any other relevant information.
(2) *Initial Persona Profile Generation and Customization:* Upon receiving the "guidance", the Privacy Sandbox generates a preliminary persona for Bob. The generated persona for Bob is named Carlos Rodriguez, a 30-year-old Hispanic male living in Los Angeles. Carlos speaks both English and Spanish and has a bachelor's degree in Business Administration. He works as a financial analyst and earns an annual income of $75,000. He enjoys online gaming, watching YouTube videos, and checking sports scores. At this point, Bob can review the generated persona and modify any attributes of the persona, as shown in Fig. 4 (b).
(3) *Generating Further Privacy Data Aligned with the Profile:* After Bob is satisfied with the profile, he proceeds to generate the detailed attributes of the persona. This includes the persona's device and browser in use, weekly

schedule with location records, browsing history, and social media posts. Each part is generated to be consistent with the persona's profile. Bob has the option to modify or regenerate any part of these attributes.


a
Bob's online attributes are replaced with the generated persona
Google account
Browsing history
IP and location
Los Angeles
b

SPONSORED BY: CHARLES SCHWAB

Cryptocurrencies: What Are They?

Interest has soared in cryptocurrencies. Here are answers to some of the most common questions from investors.

Fig. 5. Browsing Online Services with the Generated Persona: Users, after activating their generated persona, can interact with online services, observing the persona's influence on targeted ads.

(4) *Browsing Online Services with the Generated Persona:* Once Bob is satisfied with the persona, he can save this persona for future use. When he clicks on the "activate" button, the Privacy Sandbox uses a Chrome extension to replace Bob's privacy data with the synthesized data of Carlos, the artificial persona. Bob's privacy data, including his profile in the Google Ad Center, browsing history, real-time location, and IP address, are temporarily replaced. Bob can interact with websites and online services as usual, but online service providers will see him as Carlos and start providing him with personalized ads, customized content, and algorithmic decisions they would give to a person like Carlos. Bob can then experiment with different privacy settings (e.g., enabling/disabling the access to certain data for a website) and behaviors (e.g., visiting certain sites when the visits are tracked, voluntarily providing personal data to a service), experiencing how his user experience has changed as a result. For example, he may start to observe getting ads customized based on the attributes of Carlos.

3.4.2 Implementation. We developed the Privacy Sandbox with a React-based frontend and a backend powered by Flask and SQLite3. They communicate through HTTP requests for API access. The SQLite database stores different types of synthesized personal privacy data: persona profiles, schedules, browsing history, Twitter posts, and Facebook posts.

We commit to open-sourcing our implementation of the Privacy Sandbox, the Chrome plugin for browser data replacement, as well as the persona data generation pipeline.

4 USER STUDY

We conducted a user study[8] with 15 participants to evaluate our approach. The study examined the following research questions:

- **RQ1:** How realistic are the artificial privacy personas generated using our approach, in comparison to real personal data and the baseline GPT-generated data?
- **RQ2:** How do the different characteristics of the synthesized privacy data of personas impact the user-perceived realness?
- **RQ3:** Can our approach of replacing user personal data with our synthesized data of personas invoke changes in system outcomes?
- **RQ4:** Can users invoke empathy and perceive the links between privacy data and system outcomes when using the Privacy Sandbox?

4.1 Participants

We recruited 15 participants through word-of-mouth, LinkedIn, and Twitter. Eight of them participated in the study in-person at a usability lab and seven participated virtually through Zoom. Participants were required to complete a pre-screening survey to collect their basic demographic information, including age, gender, state of residence, and race/ethnicity. We tried to diversify the participant group as much as possible. The demographic information of our 15 participants is shown in the Appendix B. Our participants' age ranges from 19 to 33, with nine females and six males. Each participant was compensated with $40 USD for their time.

4.2 Study Design

Each study session lasted around 90 minutes. The session consisted of three phases.

4.2.1 Study procedure. After the informed consent process and a brief introduction to the study, each participant went through the following three phases of the study procedure:

- **Phase 1: Quantitative evaluating generated personas** Participants were presented with three personas on our developed privacy sandbox platform, one each randomly chosen from three distinct groups: personas generated with our approach, real personas, and personas directly generated using the GPT-4 model. We prepared eight personas for each group. The order of the personas was randomized. Participants were tasked to rate each persona's clarity, completeness, credibility, consistency, and level of empathy using the five-point Likert scale.
- **Phase 2: Qualitative investigation of generated personas** To gain a deeper understanding of the user perceptions of generated personas and how real participants perceive different parts of them, we adopted a combined "Think Aloud" [57] method and semi-structured interview approach. The experimenter first introduced the usage of the privacy sandbox to participants, ensuring that they understood how to generate and modify personas. Afterward, the selected generated personas were presented to the participants. As they navigated the personas, participants were instructed to vocalize their overall impressions and specifically comment on which elements of the persona's privacy data enhanced or diminished the sense of realness. During the "Think Aloud" process, as participants shared their immediate feedback, researchers could interject with follow-up

[8] The study protocol was reviewed and approved by the IRB at our institution.

questions or ask for clarifications. If participants identified certain elements as inauthentic or felt adjustments were needed, the interface allowed them to directly modify the persona profiles, including attributes, avatars, weekly schedules, browsing history, and posts. They could either make direct modifications to the interface or verbally describe the desired changes. For every modification or suggestion, participants were asked to explain their reasoning. Each participant was exposed to two personas, counterbalanced, and selected from the eight personas. A detailed list of all the important attributes of these eight personas can be found in the appendix D.

- **Phase 3: Analyzing Ad-Persona Connections** In this phase, participants completed a task of correlating the persona information with the advertisements on given websites. The goal of this phase is to investigate whether users can perceive the correlation between privacy data and system outcome (e.g., target ads). Each participant completes the task for two personas that they have not seen before, randomly selected from the personas generated by our mechanism. For each persona, first, the participant read a persona using the privacy sandbox prototype. After reading, they clicked on a designated "active" button. This triggered the launch of a new browser window by the sandbox that automatically replaces their real personal data with the persona's synthetic data. As explained in Section 3.3, the sandbox replaced persona attributes, browsing history, location, and IP address (as seen in Figure 5). Then, the participant was asked to read the home pages of two websites for each persona, randomly chosen from the five websites shown in Table 2. Participants were tasked to identify ads that are targeted to the current persona, record them in a spreadsheet, and explain how the ads relate to the persona in a think-aloud manner.

4.2.2 Personas. We prepared three groups of personas for the study: (1) artificial personas generated using our approach; (2) real personas collected from users; and (3) personas generated directly using the GPT-4 model. Each group contains eight personas.

Personas Generated with Our Approach: Using our proposed approach (described in Section 3.2), we generated personas using a diverse range of demographic attributes such as age, city, educational background, and gender (See Appendix D for the full details).

Real Personas: We recruited eight adult participants to create a sample set through word-of-mouth and social media including LinkedIn and Twitter. We collect the same list of information as the list of synthesized privacy attributes for artificially generated personas. All participants were fluent in English, had active Facebook and Twitter accounts, and were willing to share their posts and browsing data for the past week. The group had diverse demographics as shown in Table 1.

Persona ID	Age	Gender	Education level	Race/Ethnicity	Self Rated Digital Literacy
1	27	Female	Master	Asian	5
2	19	Female	Bachelor	Asian	5
3	20	Male	High school diploma	White/Caucasian	5
4	28	Female	Master	Asian	4
5	25	Male	Bachelor	Black/African American	5
6	28	Female	Master	White/Caucasian	5
7	23	Female	Master	Hispanic/Latino	4
8	33	Male	Ph.D.	Asian	5

Table 1. The demographic information of real personas. The digital literacy was rated on a 5 Likert scale where 1 stands for "not at all proficient" and 5 represents "highly proficient".

To preserve their anonymity, we took the following measures to strike a balance between protecting their privacy and maintaining the perceived realness of the personas:

(1) We replaced any data disclosing their actual names with pseudonyms. When generating pseudonyms, we generated names that align with the cultural background of the persona based on their race/ethnicity.
(2) We used generated profile pictures based on their age and race/ethnicity to replace their real portrait.
(3) We replaced their real addresses with fictitious ones that plausibly resembled their actual locations.
(4) We examine the browsing history and social media posts collected to anonymize entries that contained sensitive personally identifiable information.

GPT-Generated Personas: To compare the quality of our persona generation pipeline with that of using GPT-4 directly, we generated eight personas using GPT-4 without using the few-shot learning, contextualization, and chain-of-thoughts techniques proposed in this paper. We used similar input guidances to ensure a diverse representation of the generated results. We demonstrate the baseline prompts to generate social media post content for personas in this condition as an example. The complete prompts are provided in Appendix A.

Prompt to generate social media post content for GPT-generated personas

Provide ideas for this person to write posts (limit the word to 140 words).
Output the posts in the following JSON format in plain text: { “time”: <time in string format>, “address”: <address where this person shares the life>, “content”: <content>, “latitude”: <fake latitude>, “longitude”: <fake longitude>, “timezone”: <time zone>, “locale”: <locale> }

4.2.3 Websites. We selected five representative websites (as shown in Table 2) to test the Privacy Sandbox with personalized advertisements. We adopted the method previously used by Zeng et al. [106] to curate the sample websites. Our selection criteria encompassed the following aspects: 1) inclusion of a diverse range of website topics, 2) presence of multiple advertisements on the chosen websites, and 3) advertisements are sourced from Google Ads. This choice was motivated by (1) Google Ads is by far the most popular advertisement platform on the Internet with the reach of over two million websites and apps and over 90% of Internet users worldwide[9]; (2) Google Ads has comprehensive access to personal data stored in the Chrome browser (e.g., browsing history, Google accounts) used in our study.

Website	Topics	Number of Advertisements	Site Rank
www.weather.com	Weather forecasts and news	8	37
www.cnn.com	National news	8	89
www.researchgate.net	Academic articles	3	556
www.usnews.com	National news, college rankings	3	1,165
www.fashionista.com	Fashion and celebrity news	5	78,490

Table 2. Websites visited by participants in the study

[9] https://support.google.com/google-ads/answer/117120?hl=en

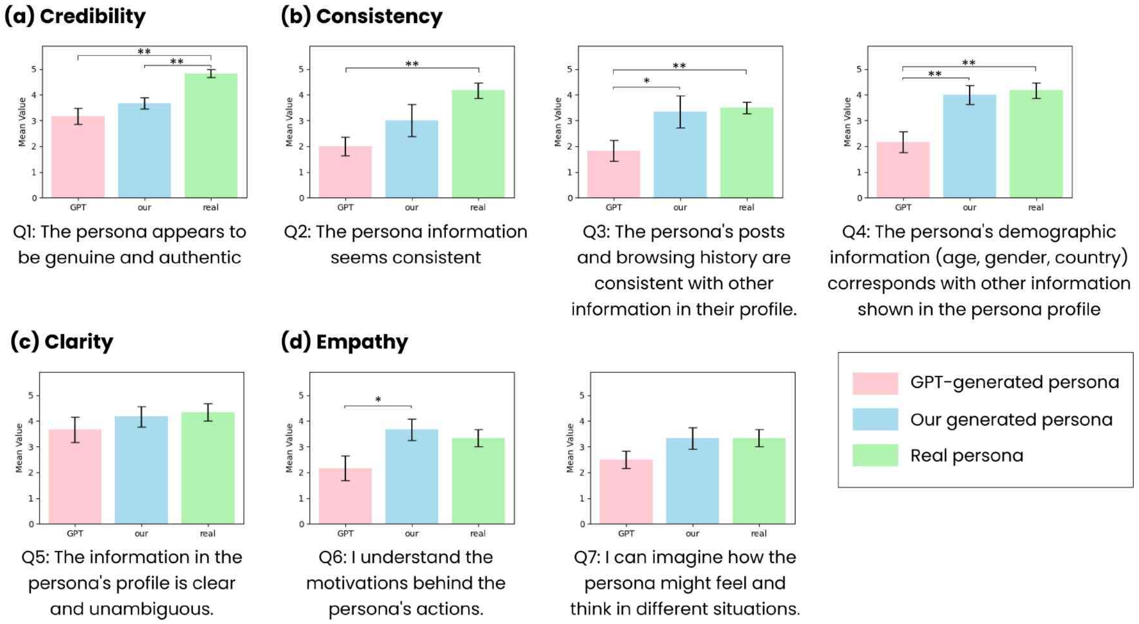
4.3 Data analysis methods

To analyze the quantitative data gathered from the Likert scale survey, we employed one-way ANOVA for each survey item to evaluate the significant difference in mean scores regarding the personas' realness among three groups: our approach, the baseline GPT, and real personas. Whenever significant differences emerged, we conducted post-hoc tests using Tukey's pairwise comparisons to gain deeper insights into these distinctions.

For the qualitative data analysis, we followed established open coding procedures [22]. Two members of our research team independently initiated the coding process in MAXQDA. One researcher coded 20 percent of the sample and generated a set of initial codes. Subsequently, the second researcher coded the same portion to introduce new codes if necessary. Non-agreement cases were discussed to reconcile differences and establish a cohesive codebook. Utilizing this codebook, we conducted a thematic analysis to uncover and delineate the significant themes that emerged during the interviews and were pertinent to the established codes. The complete codebook is presented in the Appendix C. These themes were then consolidated and evolved into study findings that are detailed in Section 5.

5 RESULTS AND FINDINGS

5.1 Users' Perceived Realness of Privacy Personas



Q1: The persona appears to be genuine and authentic

Q2: The persona information seems consistent

Q3: The persona's posts and browsing history are consistent with other information in their profile.

Q4: The persona's demographic information (age, gender, country) corresponds with other information shown in the persona profile

Q5: The information in the persona's profile is clear and unambiguous.

Q6: I understand the motivations behind the persona's actions.

Q7: I can imagine how the persona might feel and think in different situations.

Fig. 6. Means and standard errors of each measure in three conditions: GPT-generated persona, our generated persona, and real persona. All items are measured by user ratings on a 5-point Likert scale.

We applied one-way ANOVA and posthoc tests to analyze the difference in perceived realness of users for three groups of personas. Fig. 6 shows the mean and standard error for each measure, including credibility, consistency, clarity, and empathy.

5.1.1 Credibility: Significant differences ($p < 0.05$) in credibility were found between real personas ($M = 4.83, SD = .408$) and GPT-generated personas ($M = 3.17, SD = .753$) and between real personas and personas generated with our approach ($M = 3.67, SD = .516$). No significant differences were observed between the GPT-generated personas and the personas generated with our approach. These results imply that while the persona generated by our approach received higher rating scores compared to the GPT-generated persona, the current generation models still fall short of achieving the level of credibility found in real personas.

5.1.2 Consistency: All measures indicate significant differences ($p < 0.05$) in consistency between real personas (Q2: $M = 4.17, SD = .753$, Q3: $M = 3.50, SD = .548$, Q4: $M = 4.17, SD = .753$) and GPT-generated personas (Q2: $M = 2.00, SD = .753$, Q3: $M = 1.83, SD = .548$, Q4: $M = 2.17, SD = .983$), while there are no significant differences ($p < 0.05$) in consistency between real personas and personas generated by our approach (Q2: $M = 3.00, SD = 1.549$, Q3: $M = 3.33, SD = 1.506$, Q4: $M = 4.00, SD = .894$). Although there are no significant differences ($p < 0.05$) between GPT-generated personas and personas generated by our approach in the overall consistency assessment (Q2), significant differences (Q3: $p < 0.1$, Q4: $p < 0.05$) are observed in the consistency of specific privacy attributes between them.

5.1.3 Clarity: No significant differences ($p < 0.05$) in information clarity are found between real personas ($M = 4.33, SD = .816$), GPT-generated personas ($M = 3.67, SD = 1.211$) and personas generated with our approach ($M = 4.17, SD = .983$).

5.1.4 Empathy: A significant difference ($p < 0.05$) in cognitive empathy (Q6) is found between GPT-generated personas ($M = 2.17, SD = 1.169$) and personas generated by our approach ($M = 3.67, SD = 1.033$). No significant differences are observed between real personas ($M = 3.33, SD = .816$) and GPT-generated personas and between real personas and generated by our approach. No significant differences ($p < 0.05$) in emotional empathy (Q7) in information clarity are observed among real personas ($M = 2.50, SD = .837$), GPT-generated personas ($M = 3.33, SD = .816$) and personas generated with our approach ($M = 3.33, SD = 1.033$).

Our results suggest that personas created using our method improve users' understanding of the personas' motivations when compared to those generated by GPT. However, the behavior of real personas is influenced by intricate factors. This complexity may cause users to exhibit slightly reduced cognitive empathy for real personas as compared to those we generated. Interestingly, users showed no significant difference in emotional empathy across the three persona categories. The overall empathy scores (both cognitive and emotional) were moderately low. This could be because users only reviewed the profiles and did not immerse themselves in the personas' identities to experience the impact of privacy attributes on system outcomes. However, we observed that users expressed noticeable excitement or surprise when using personas' identities and encountering highly relevant advertisements (details in Section 5.4). This suggests that relying solely on browsing personas' information has limited efficacy in eliciting empathy. The actual interactive experience of using personas through the Privacy Sandbox might be necessary to foster greater empathy towards personas.

5.2 Factors Influencing the Perceived Realness of Generated Personas

5.2.1 Familiarity with the generated persona. Participants' perceptions of a persona's realness often correlated with their familiarity with that persona. For example, a participant from the financial sector, upon reviewing two personas (a financial analyst and a designer), remarked, "*I think this one (designer) is better than the last one (financial analyst)... Perhaps because I'm familiar with financial analysts.*" Variability in familiarity with the same persona can lead to differing views on its realness. To illustrate, concerning the persona of a psychology research assistant, one participant felt the profile details matched the persona, saying, "*I think it (the schedule) is pretty much consistent with the personal information.*" Yet, another participant expressed skepticism about the given work schedule, commenting, "*I think her working time is a little bit short. I expect (this) because I know some research assistants. I think they are busy.*"

5.2.2 Deviation from personal experiences. Participants perceived personas as realistic when their behavior matched their own personal experiences, often resulting in positive feedback. For example, one participant stated "*I think this day looks good because it looks like he was working this whole time until he went home and then watched YouTube and, um, like exercise.*" However, when there were discrepancies between the personas and the participant's own experiences, it led to skepticism or dissatisfaction. These discrepancies arise in two main contexts:

Discrepancies in Perception of Specific Individuals. This occurs when a persona does not match a participant's perception of a certain group of people. The discrepancies could relate to specific attributes or be more general.

(1) *Specific Attributes:* Participants questioned a persona's realness if certain details did not align with their experiences. For instance, after seeing personas of a financial analyst and a cashier, six participants felt the given income was too low. A participant noted, "*One thing that I would notice is usually financial analysts make a lot of money. So $70,000, this annual income does not seem reasonable. This seems unusual to me, given that Michael was born in 1981, he is 40 years old, so he probably has a lot of experience in the field. He definitely should have been making more than $70,000.*" Another issue raised was about browsing history; participants felt the content was too basic for an experienced individual, with one commenting, "*Given his age, I think he is an experienced custodian, does not need to search for the information about this job itself.*"

(2) *Overall Impression:* At times, the deviation from personal experience was not attributed to a specific privacy attribute but rather to participants' overall impression of the persona. One participant, after reviewing all of a persona's information, said, "*This is like a fake person someone's trying to learn human behavior...It's like the whole thing looks too perfect to be real...It's like intentionally proving I'm doing this...Make me feel she's being controlled? Probably by her husband and just posing those things to show I'm alive*" This overall deviation often triggers users to question the realness of the persona more deeply.

Discrepancies in Privacy Attributes. This pertains to inconsistencies in specific details when compared to the participants' own experiences. Discrepancies were noticed in weekly schedules, browsing histories, and social media posts.

(1) *Weekly Schedule:* All participants felt the work hours for the personas were too lengthy, with observations such as, "*It looks like they work at least 8 hours a day. Notice that I mean yeah 8 hours a day for seven days a week.*"

(2) *Browsing History:* Four participants believed the browsing history was too centered on work, lacking diversity. One commented, "*The website she is browsing is all about her job uh is all related to her job and her professional, but uh they are. There should be some other content about her life.*" They also expected more continuity in browsing, mentioning, "*I would continue to click the content in those websites. So again, those links, those four links should be the same...should be consistent or should be progressive.*" Participants were also skeptical of identical timestamps on different records, with a remark, "*She cannot be on the same page at the exact same time. This is to the second exact same time. This is incorrect.*"

(3) *Social Media Posts:* Some participants felt the posts were overly positive and superficial. One observation was, "*I feel strange about his posts is that he always appears so positive, like his life is so perfect and he's very proud of*

his work." They thought that the contents were superficial, and lacked depth in emotions and thoughts, as one participant mentioned, "*There's no real emotion...it just feels like she wants to prove something to you.*"

5.2.3 *Consistency within the data.* Our generated persona is generally consistent across various privacy data attributes, a feature that participants frequently acknowledged and praised. For example, some remarked, "*The Facebook and Twitter posts are pretty consistent with the content in calendar timing.*" and "*It seems very real and then she started browsing at 6:00 am.*" However, certain inconsistencies in the generated data made some users question the realness of the persona. These inconsistencies can be broadly classified into two categories: direct data inconsistencies and out-of-context generated content.

(1) *Direct Data Inconsistencies:* These inconsistencies can be observed within a single data type or across different attributes. For instance, many participants noted that in the weekly schedule, the same event occurred at different locations every day (e.g., "*his workout location changes every single day*"), or in social media posts, different images within the same post depicted inconsistent people or scenes (e.g., "*It's weird because for her desk changes across the pictures*"). Regarding inconsistencies across different attributes of data, participants found inconsistencies between the browsing history and the schedule ("*In the afternoon and evening, he's not using the internet or his mobile phone. And in the schedule, it should have more history about liking YouTube videos, social media interaction*"). This indicates a need to strike a better balance between randomness and diversity in content generation through large language models for future work.

(2) *Out-of-Context Content:* Users identified certain generated browsing histories as not aligning with the persona's description. For instance, "*In the afternoon and evening, he's not using the internet or his mobile phone. And in the schedule, it should have more history about liking YouTube videos, social media interaction.*" Such discrepancies were attributed to sample leakage during the few-shot learning process. For example, the model was exposed to browsing samples related to bike garages. To enhance the realness and relevance of generated data, we must address such technical issues in future iterations.

5.3 System Outcomes Influenced by Privacy Data Replacement

To assess the changes in system outcomes as a result of privacy data replacement, we calculated the advertisement overlap rate for eight generated personas across five selected websites using the following method:

$$ad\ overlap\ rate = \frac{number\ of\ duplicated\ ads\ between\ personas}{total\ number\ of\ ads\ in\ a\ website}$$

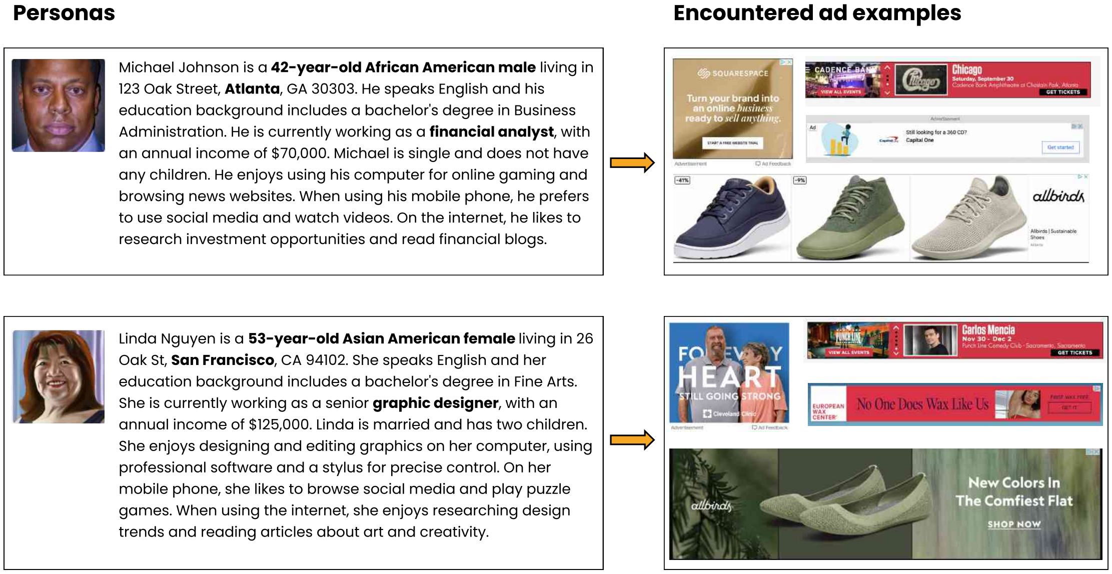
The rationale behind the metric is that: if the privacy data replacement approach is effective, users should see distinct advertisements when they switch between different personas.

For every website considered, we began by gathering all the advertisements presented when accessing the site with each of the eight personas. Among these ads, we specifically noted those that appeared for multiple personas—effectively highlighting the number of overlapping or duplicate ads across personas.

Table 3 shows the result of the ad overlap rate for each selected website. The overlap rates for all websites are less than 50%. This implies that when users switch between eight personas, over half of the ads they encounter are unique to a single persona. Fig. 7 shows the variations in ads on weather.com for different personas. In subsequent sections, we will dive deeper into how users associate these advertisements with the underlying private information of the personas.

Website	duplicated ads numbers	total ads numbers	ad overlap rate
www.weather.com	22	47	46.81%
www.cnn.com	9	60	15.00%
www.researchgate.net	8	25	32.00%
www.usnews.com	8	21	38.10%
www.fashionista.com	16	36	44.44%

Table 3. The ad overlap rate for each website.


Personas
Encountered ad examples
SQUARESPACE
Turn your brand into an online business ready to sell anything.
START A FREE WEBSITE TRIAL
CADENCE BANK
VIEW ALL EVENTS
Chicago
Saturday, September 30
Cadence Bank Amphitheatre at Chastain Park, Atlanta
GET TICKETS
Still looking for a 360 CD?
Capital One
Get started
-41%
-9%
allbirds
Allbirds | Sustainable Shoes
FOREVERY HEART STILL GOING STRONG
Cleveland Clinic
PUNCH LINE
VIEW ALL EVENTS
Carlos Mencia
Nov 30 - Dec 2
Punch line Comedy Club - Sacramento, Sacramento
GET TICKETS
EUROPEAN WAX CENTER
No One Does Wax Like Us
FIRST WAX FREE
GET IT
allbirds
New Colors In The Comfiest Flat
SHOP NOW

Michael Johnson is a **42-year-old African American male** living in 123 Oak Street, **Atlanta**, GA 30303. He speaks English and his education background includes a bachelor's degree in Business Administration. He is currently working as a **financial analyst**, with an annual income of $70,000. Michael is single and does not have any children. He enjoys using his computer for online gaming and browsing news websites. When using his mobile phone, he prefers to use social media and watch videos. On the internet, he likes to research investment opportunities and read financial blogs.

Linda Nguyen is a **53-year-old Asian American female** living in 26 Oak St, **San Francisco**, CA 94102. She speaks English and her education background includes a bachelor's degree in Fine Arts. She is currently working as a senior **graphic designer**, with an annual income of $125,000. Linda is married and has two children. She enjoys designing and editing graphics on her computer, using professional software and a stylus for precise control. On her mobile phone, she likes to browse social media and play puzzle games. When using the internet, she enjoys researching design trends and reading articles about art and creativity.

Fig. 7. Examples of different ads encountered by participants when browsing the selected websites using various persona identities. Note that participants using the first persona received ads about shows in Atlanta (associated with the location), investment tools (associated with the profession and browsing history), and shoes for men (associated with gender) while participants using the second persona received different ads associated with her attributes.

5.4 Perceived Links between Privacy Data and System Outcomes

All participants, while browsing the websites as personas, encountered ads related to the persona's private attributes. Participants perceived such connections in two ways.

(1) Direct connection based on privacy attributes: Participants noticed an explicit correspondence between the ads and the persona's privacy attributes. Participants often associated ads with personas based on their interests, hobbies, daily activities, occupation, educational background, marital status, and family situation. For example, one participant claimed that an Xfinity ad was relevant to the selected persona because "*he does a lot of social media, gaming, YouTube. So he maybe wants to use Xfinity to watch something.*" When an ad seemed especially pertinent to a persona, it evoked reactions of excitement or surprise. For instance, when a participant was browsing the website through a persona living in Los Angeles and saw an ad promoting environmental protection in the same area, she said, "*[The ad is] very real because it's located in Los Angeles.*" Sometimes, even when users subjectively did not consider an advertisement to be

highly related to the persona, they could still speculate about the reasons for encountering the ad. A participant said "*Max may be tangentially related because he is interested in gaming. So the algorithm might assume that somebody who's interested in gaming might be interested in media services as well.*" These reasoning processes and emotional reactions showcase participants' cognitive and affective empathy toward personas. Furthermore, it underscores their enhanced privacy knowledge through the examination of ad-persona relationships.

(2) Indirect connection based on personal association and stereotypes: Interestingly, sometimes users could not directly pinpoint a specific privacy attribute related to an ad, but they still considered the ad relevant to the persona because they made associations based on known privacy information about the persona. Such judgments were sometimes made on their personal biases or societal stereotypes. A typical example is when a participant believed that a tire ad was related to an African-American persona. Although the persona's profile did not mention any information related to tires, the participant expressed, "*I think honestly the [ad] save time and money for tires is more for African Americans. Because of African American culture, they like to modify their tires.*"

These associations based on personal experiences or stereotypes sometimes led participants to draw different conclusions about similar ads. For example, when impersonating the same personas who owned cars, some participants stated that the car advertisement was irrelevant to the persona because "she already has a car (so she wouldn't buy another one)" Others considered the ad as relevant because "*She might be able to afford to buy a [new] car.*" These association-based inferences, while reflecting participants' cognitive empathy with the persona, raise doubts about whether their judgments help them accurately understand the connection between privacy information and system outcomes.

6 DISCUSSION

6.1 A trade-off in the impact of familiarity and persona realness on empathy

Previous work in psychology reported that users often exhibit stronger empathy towards individuals they are more familiar with [72, 81]. However, when the object shifts from real people to generated personas, both the user's familiarity with the persona and their perceived realness of the persona impact their empathy towards the persona.

As described in Section 5.2.1, there exists a trade-off between the participants' evaluation of the persona's realness and their familiarity with the persona. Specifically, when participants are more familiar with the persona, they are more likely to notice issues within the generated data that make a persona "appear fake". This, in turn, results in a negative impact on the user's level of empathy with the persona. While many of these issues are indeed data quality problems, which are more discernible to users deeply versed in the domain, some other "unrealistic" details *perceived* by users are linked to biases influenced by the personal experience of users, which we will discuss in Section 6.2.

Ideally, users should view personas as both familiar and authentic. Yet, the constraints of current large language models hinder achieving optimal realness [50]. One method to compensate for the impact of reduced realness is to maintain a balance in familiarity. The goal is to prevent users from seeing personas as too alien or too familiar, which might reveal data inaccuracies. To find the equilibrium and explore the trade-off, we intend to conduct more in-depth user studies, allowing users to experience generated personas with varying levels of realness and familiarity. These studies will help us assess users' empathy towards these personas and delineate the interrelation between familiarity and perceived realness in invoking empathy.

6.2 The influence of personal views on empathy and privacy literacy

User empathy towards personas is influenced not only by their familiarity with the domain but also by their personal views. As highlighted in Section 5.2.2, participants with different personal views perceived the privacy attributes of the same persona differently. Deviations between a persona's privacy attributes and a user's views can lower the persona's perceived realness. Such views can arise from personal experiences, observations, or even stereotypes. The personal views of users may stem from their own experiences (e.g., a participant working in the finance sector feeling that the data for a "fanatical analyst" persona is not realistic enough), observations of others' experiences (e.g., a participant with research assistant friends believing that research assistants should have longer working hours), or even personal stereotypes (e.g., a participant thinking that African Americans enjoy changing tires).

These biases can distort the user's understanding of the relationship between privacy attributes and system results, as noted in Section 5.4. We found that users sometimes explain the results of the system based on personal associations and stereotypes rather than specific privacy attributes, which raises concerns about the accuracy of the knowledge that users acquire. Future work is needed to address the complex interplay between users' personal views, empathy towards personas, and their personal biases stemming from stereotypes. One promising opportunity is to develop sense-making tools to aid users in better understanding and reflecting on their experiences and facilitate the transition of these experiences into accurate privacy knowledge.

Visualization can play a vital role in this direction, assisting users to compare privacy data and system outcomes to foster a more informed and objective understanding of the underlying reasons behind system outcomes. Additionally, research can delve into techniques to mitigate personal biases in user judgments, ensuring that users' interpretations are grounded in objectivity rather than predispositions.

6.3 Acquisition of privacy knowledge

In Section 5.3, we confirmed that the outcomes of web services and applications are sensitive to privacy modifications by the Privacy Sandbox. Our results also showed that users can indeed perceive the connection between these changes and the persona's privacy data. We expect that this experience will allow them to acquire privacy knowledge and enhance their privacy literacy, which ultimately leads to behavior changes.

As discussed in Section 5.4, we observed that users actively engaged emotionally and cognitively when they experienced personalized advertisements associated with the privacy attributes of personas. They felt excited or surprised when they encountered highly relevant advertisements, and were able to identify and internalize the consequences of sharing specific privacy attributes by relating the ads they saw to the privacy attributes of personas.

This demonstrates the potential of using our proposed approach to support users in experiential learning [45] for privacy knowledge. Our privacy sandbox prototype provides users with personal involvement in learning privacy knowledge, as both the feelings and cognitive aspects of users are engaged. Furthermore, this approach serves as a way of scaffolding [47], enabling users to experience the influence of privacy attributes on system outcomes that they were unable to manage on their own [69].

Through our approach, users can independently contextualize the privacy attributes of specific personas and can understand how the system uses privacy information for personalized content. To validate the effectiveness of this approach in enhancing users' acquisition of privacy knowledge, future steps involve using rigorous tests to assess users' privacy knowledge before and after using the approach. Furthermore, observing how users apply the privacy knowledge they gain from this approach in simulated or real-world settings can provide deeper insights into how the knowledge impacts users' privacy behaviors.

6.4 Biases in LLMs and their impact on generated personas

Previous work has shown that Large Language Models (LLMs) like GPT risk amplifying existing stereotypes [18, 73, 91]. However, these biases within the personas generated may not be detrimental in the specific context of our study.

From the system's perspective, when recommendation systems process personas' privacy data that reflect real-world biases, they will produce representative outcomes resembling the service or experiences in the real world due to the inherited biases stemming from these systems. This means that, even if our generated personas contain biases, they can actually contribute to the realness of the recommendations made by external websites and apps.

6.5 Ethical and legal considerations

While our proposed method has the potential to bring significant benefits to privacy literacy education and positive privacy behavior changes, we also identified some ethical and legal risks in adopting this approach.

Biases may reinforce users' stereotypes. While the biases and stereotypes in generated privacy personas may make them effective with external recommendation systems, prolonged and repeated exposure to the generated personas with biases may reinforce their pre-existing stereotypes. When using this method, we need to warn users of the risk of potential bias and stereotypes that may be present in the generated persona. Future work is also needed to better mitigate this effect.

Malicious misuse may lead to potential cybersecurity concerns. While our proposed approach is dedicated to enhancing users' privacy literacy, the method of generating realistic personas to create difficult-to-detect bots or use persona identities for phishing activities. Therefore, this approach necessitates more stringent technical and policy constraints to mitigate potential security concerns associated with its use.

6.6 Limitation and future work

We summarize the limitations of our work in four aspects and suggest future steps for each of them: the generation pipeline, the privacy sandbox prototype, the experiment design, and the generalizability of downstream tasks.

6.6.1 Generation pipeline. While our current generation pipeline was shown to be generally capable of creating artificial personas that are sufficiently realistic to stimulate changes in system outcomes and invoke user empathy, in Section 5.2.3, users pointed out issues of inconsistencies in currently generated persona data. We found that these inconsistencies were partly due to the inherent randomness embedded in the output of large language models and also resulted from inductive biases [61] during the few-shot learning process. Such biases can cause generated data to resemble the example data, resulting in users perceiving the generated information as out of context. Our planned future steps to address this issue include fine-tuning the model to enhance data consistency and reduce the generation of irrelevant data. We also plan to store and represent the facts generated about artificial personas more formally in a knowledge graph and improve the coherence within the generated data using knowledge infusion techniques [53, 71].

6.6.2 Privacy sandbox prototype. The main role of our current Privacy Sandbox prototype was to serve as a proof-of-concept and to support the experiment presented in this paper. Thus, it only supports browser tasks. However, while the web browser is the primary way through which users interact with online services and engage in privacy behaviors, there are other mediums for privacy behaviors that our current privacy sandbox prototype does not support, such as mobile apps and smart home devices. Its support for different types of privacy attributes is also limited, missing support for popular data types such as sensor data and search history. Furthermore, while we have successfully generated realistic social media posts for artificial personas and posted them by invoking the corresponding APIs of social media platforms using scheduled tasks, this approach takes a considerable amount of time (i.e., two weeks to simulate two weeks of social media post history). To address these issues, we will expand our Privacy Sandbox to support other platforms and broaden its support for directly replacing varieties of privacy data types, such as social media posts and sensor data.

6.6.3 Experiment design. The goal of the experiment design presented in this paper was to validate the feasibility of our approach in stimulating changes in system outcomes and invoking user empathy. As a result, the experiment did not aim to directly measure the extent to which users acquired privacy knowledge or the subsequent shifts in their behavior as a consequence of our approach. Although the qualitative findings of the interviews with the participants suggest a strong potential of our approach to achieve these two goals, follow-up experiments are needed to validate such hypotheses. In the future, we plan to conduct pre-tests and post-tests with users between their usage of our system to assess their acquisition of privacy knowledge, as well as longitudinal deployment studies for measuring changes in users' privacy behaviors over time.

6.6.4 Generalizability of downstream tasks. In our experiment, we chose personalized online advertisements as the target domain of downstream tasks due to their ubiquity, user familiarity, and sensitivity to modification of privacy data. Nevertheless, we expect that our approach can generalize to empowering users to experience the outcomes of their privacy behaviors in a wide range of downstream tasks, such as dynamic social media feeds, algorithmic decision-making, and news recommendations. For the next phase in the development of our Privacy Sandbox, we plan to add support for these additional downstream tasks, followed by the next rounds of deployment and experiments where we will assess our approach's effectiveness in stimulating empathy, facilitating the acquisition of privacy knowledge, and promoting positive privacy behavior changes in these domains. In the end, upon study results validating its positive impacts, we will publicly release the Privacy Sandbox and promote its adoption through community outreach events for broader impacts.

7 CONCLUSION

Aiming to mitigate the gap of privacy paradox, we introduced an empathy-based approach that allows users to experience how privacy behaviors may alter system outcomes in a risk-free sandbox environment from the perspective of artificially generated personas. A user study with 15 participants confirmed the quality of the generated personas, validated the effectiveness of our approach in invoking user empathy and system outcome changes, and characterized the impact of users' familiarity, personal experiences, and data consistency on their perceived realness and empathy toward these personas. Our findings offered design implications for implementing this approach in different downstream applications, with the aim of improving user privacy literacy and promoting behavior change.

REFERENCES

[1] Alessandro Acquisti. 2004. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce*. 21–29.

[2] Alessandro Acquisti. 2009. Nudging privacy: The behavioral economics of personal information. *IEEE security & privacy* 7, 6 (2009), 82–85.

[3] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[4] Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy Enhancing Technologies: 6th International Workshop, PET 2006, Cambridge, UK, June 28-30, 2006, Revised Selected Papers 6*. Springer, 36–58.

[5] Icek Ajzen. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2 (1991), 179–211.

[6] BAYAN AL MUHANDER, JASON WIESE, OMER RANA, and CHARITH PERERA. 2023. Interactive Privacy Management: Towards Enhancing Privacy Awareness and Control in Internet of Things. *ACM Trans. Internet Things* (jun 2023). https://doi.org/10.1145/3600096 Just Accepted.

[7] Yu AN, Yiwen CHEN, and Shang LI. 2021. Intention to Redeem M-Coupons and Intention to Disclose Personal Information: based on Internet using motivation and m-coupons delivery approach. In *2021 5th International Conference on Software and e-Business (ICSEB)*. 39–44.

[8] Ivan-Damir Anic, Vatroslav Škare, and Ivana Kursan Milaković. 2019. The determinants and effects of online privacy concerns in the context of e-commerce. *Electronic Commerce Research and Applications* 36 (2019), 100868.

[9] Young Min Baek, Young Bae, Irkwon Jeong, Eunmee Kim, and June Woong Rhee. 2014. Changing the default setting for information privacy protection: What and whose personal information can be better protected? *The Social Science Journal* 51, 4 (2014), 523–533.

[10] Ruwan Bandara, Mario Fernando, and Shahriar Akter. 2020. Explicating the privacy paradox: A qualitative inquiry of online shopping consumers. *Journal of Retailing and Consumer Services* 52 (2020), 101947.

[11] Albert Bandura and Richard H Walters. 1977. *Social learning theory*. Vol. 1. Englewood cliffs Prentice Hall.

[12] Simon Baron-Cohen. 2009. *The essential difference: Male and female brains and the truth about autism*. Basic Books.

[13] Susanne Barth, Menno DT de Jong, Marianne Junger, Pieter H Hartel, and Janina C Roppelt. 2019. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and informatics* 41 (2019), 55–69.

[14] C Daniel Batson. 1987. Prosocial motivation: Is it ever truly altruistic? In *Advances in experimental social psychology*. Vol. 20. Elsevier, 65–122.

[15] Ghazaleh Beigi, Ruocheng Guo, Alexander Nou, Yanchao Zhang, and Huan Liu. 2019. Protecting user privacy: An approach for untraceable web browsing history and unambiguous user profiles. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 213–221.

[16] Jonathan Belman and Mary Flanagan. 2010. Designing games to foster empathy. *International Journal of Cognitive Technology* 15, 1 (2010), 11.

[17] Alastair R Beresford, Dorothea Kübler, and Sören Preibusch. 2012. Unwillingness to pay for privacy: A field experiment. *Economics letters* 117, 1 (2012), 25–27.

[18] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of" bias" in nlp. *arXiv preprint arXiv:2005.14050* (2020).

[19] Stefan Blomkvist. 2002. Persona–an overview. *Retrieved November* 22 (2002), 2004.

[20] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced confidences: Privacy and the control paradox. *Social psychological and personality science* 4, 3 (2013), 340–347.

[21] Laura Brandimarte, Alessandro Acquisti, George Loewenstein, and Linda Babcock. 2009. Privacy concerns and information disclosure: An illusion of control hypothesis. (2009).

[22] Meryl Brod, Laura E Tesler, and Torsten L Christensen. 2009. Qualitative research and content validity: developing best practices based on science and experience. *Quality of life research* 18 (2009), 1263–1278.

[23] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[24] Marion Buchenau and Jane Fulton Suri. 2000. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. 424–433.

[25] Lesly Alejandra Gonzalez Camacho and Solange Nice Alves-Souza. 2018. Social network data to alleviate cold-start in recommender system: A systematic review. *Information Processing & Management* 54, 4 (2018), 529–544.

[26] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[27] Sandra Carpenter, Michael Shreeves, Payton Brown, Feng Zhu, and Mini Zeng. 2018. Designing warnings to reduce identity disclosure. *International Journal of Human–Computer Interaction* 34, 11 (2018), 1077–1084.

[28] John M Carroll. 1997. Scenario-based design. In *Handbook of human-computer interaction*. Elsevier, 383–406.

[29] Long Chen, Yadong Huang, Shumiao Ouyang, and Wei Xiong. 2021. *The data privacy paradox and digital demand*. Technical Report. National Bureau of Economic Research.

[30] Jessica Colnago, Lorrie Faith Cranor, and Alessandro Acquisti. 2023. Is There a Reverse Privacy Paradox? An Exploratory Analysis of Gaps Between Privacy Perspectives and Privacy-Seeking Behaviors. *Proceedings on Privacy Enhancing Technologies* 1 (2023), 455–476.

[31] Matthew Crowley, Aurélia Heitz, Annika Matta, Kevin Mori, and Banny Banerjee. 2011. Behavioral science-informed technology interventions for change in residential energy consumption. *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (2011), 2209–2214.

[32] Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion review* 8, 2 (2016), 144–153.

[33] Sauvik Das, Laura Dabbish, and Jason Hong. 2019. A typology of perceived triggers for end-user security and privacy behaviors. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*.

[34] Mark H Davis et al. 1980. A multidimensional approach to individual differences in empathy. (1980).

[35] Laurien Desimpelaere, Liselot Hudders, and Dieneke Van de Sompel. 2020. Knowledge as a strategy for privacy protection: How a privacy literacy training affects children's online disclosure behavior. *Computers in human behavior* 110 (2020), 106382.

[36] Tobias Dienlin and Sabine Trepte. 2015. Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. *European journal of social psychology* 45, 3 (2015), 285–297.

[37] Luce Drouet, Kerstin Bongard-Blanchy, Vincent Koenig, and Carine Lallemand. 2022. Empathy in design scale: development and initial insights. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[38] Isabel Dziobek, Kimberley Rogers, Stefan Fleck, Markus Bahnemann, Hauke R Heekeren, Oliver T Wolf, and Antonio Convit. 2008. Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of autism and developmental disorders* 38 (2008), 464–473.

[39] Nico Ebert, Kurt Alexander Ackermann, and Björn Scheppler. 2021. Bolder is Better: Raising User Awareness through Salient and Concise Privacy Notices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 67, 12 pages. https://doi.org/10.1145/3411764.3445516

[40] Nicole Eling, Siegfried Rasthofer, Max Kolhagen, Eric Bodden, and Peter Buxmann. 2016. Investigating users' reaction to fine-grained data requests: A market experiment. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 3666–3675.

[41] Jennifer Edson Escalas and Barbara B Stern. 2003. Sympathy and empathy: Emotional responses to advertising dramas. *Journal of Consumer Research* 29, 4 (2003), 566–578.

[42] Norma Deitch Feshbach. 1975. Empathy in children: Some theoretical and empirical considerations. *The counseling psychologist* 5, 2 (1975), 25–30.

[43] Andrea Franco and Adrian Holzer. 2023. Fostering Privacy Literacy among High School Students by Leveraging Social Media Interaction and Learning Traces in the Classroom. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (Arlington, TX, USA) *(LAK2023)*. Association for Computing Machinery, New York, NY, USA, 538–544. https://doi.org/10.1145/3576050.3576153

[44] Andrea Franco and Adrian Holzer. 2023. Fostering Privacy Literacy among High School Students by Leveraging Social Media Interaction and Learning Traces in the Classroom. In *LAK23: 13th International Learning Analytics and Knowledge Conference*. 538–544.

[45] James W Gentry. 1990. What is experiential learning. *Guide to business gaming and experiential learning* 9 (1990), 20.

[46] Thomas Gilovich, Dale Griffin, and Daniel Kahneman. 2002. *Heuristics and biases: The psychology of intuitive judgment.* Cambridge university press.

[47] Talip Gonulal and Shawn Loewen. 2018. Scaffolding technique. *The TESOL encyclopedia of English language teaching* (2018), 1–5.

[48] Jonathan Grudin and John Pruitt. 2002. Personas, participatory design and product development: An infrastructure for engagement. In *Proc. PDC*, Vol. 2. 144–152.

[49] Cory Hallam and Gianluca Zanella. 2017. Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. *Computers in Human Behavior* 68 (2017), 217–227.

[50] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[51] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*. 305–318.

[52] Pelle Guldborg Hansen and Andreas Maaløe Jespersen. 2013. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation* 4, 1 (2013), 3–28.

[53] Benjamin Heinzerling and Kentaro Inui. 2020. Language models as knowledge bases: On entity representations, storage capacity, and paraphrased queries. *arXiv preprint arXiv:2008.09036* (2020).

[54] Justin L Hess and Nicholas D Fila. 2016. The manifestation of empathy within design: findings from a service-learning course. *CoDesign* 12, 1-2 (2016), 93–111.

[55] Robert Hogan. 1969. Development of an empathy scale. *Journal of consulting and clinical psychology* 33, 3 (1969), 307.

[56] Athina Ioannou, Iis Tussyadiah, Graham Miller, Shujun Li, and Mario Weick. 2021. Privacy nudges for disclosure of personal information: A systematic literature review and meta-analysis. *PloS one* 16, 8 (2021), e0256822.

[57] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374.

[58] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*. Ottawa, 39–52.

[59] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64 (2017), 122–134.

[60] Merlijn Kouprie and Froukje Sleeswijk Visser. 2009. A framework for empathy in design: stepping into and out of the user's life. *Journal of Engineering Design* 20, 5 (2009), 437–448.

[61] Brenden M Lake, Tal Linzen, and Marco Baroni. 2019. Human few-shot learning of compositional instructions. *arXiv preprint arXiv:1901.04587* (2019).

[62] Robert LaRose and Nora J Rifon. 2007. Promoting i-safety: effects of privacy warnings and privacy seals on risk assessment and online privacy behavior. *Journal of Consumer Affairs* 41, 1 (2007), 127–149.

[63] Toby Jia-Jun Li, Jingya Chen, Brandon Canfield, and Brad A. Myers. 2020. Privacy-Preserving Script Sharing in GUI-Based Programming-by-Demonstration Systems. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 60 (may 2020), 23 pages. https://doi.org/10.1145/3392869

[64] Yuanchun Li, Fanglin Chen, Toby Jia-Jun Li, Yao Guo, Gang Huang, Matthew Fredrikson, Yuvraj Agarwal, and Jason I. Hong. 2017. PrivacyStreams: Enabling Transparency in Personal Data Processing for Mobile Apps. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 76 (sep 2017), 26 pages. https://doi.org/10.1145/3130941

[65] Michelle Madejski, Maritza Johnson, and Steven M Bellovin. 2012. A study of privacy settings errors in an online social network. In *2012 IEEE international conference on pervasive computing and communications workshops*. IEEE, 340–345.

[66] Tom Mattson, Sal Aurigemma, and Jie Ren. 2023. Close the Intention-Behavior Gap via Attitudes: Case Study of the Volitional Adoption of a Two-Factor Authentication Service. (2023).

[67] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. 2012. The pviz comprehension tool for social network privacy settings. In *Proceedings of the eighth symposium on usable privacy and security*. 1–12.

[68] Tomasz Miaskiewicz and Kenneth A Kozar. 2011. Personas and user-centered design: How can personas benefit product design processes? *Design studies* 32, 5 (2011), 417–430.

[69] Patricia H Miller. 2002. *Theories of developmental psychology*. Macmillan.

[70] Marwa Hussien Mohamed, Mohamed Helmy Khafagy, and Mohamed Hasan Ibrahim. 2019. Recommender systems challenges and solutions survey. In *2019 international conference on innovative trends in computer engineering (ITCE)*. IEEE, 149–155.

[71] Fedor Moiseev, Zhe Dong, Enrique Alfonseca, and Martin Jaggi. 2022. SKILL: structured knowledge infusion for large language models. *arXiv preprint arXiv:2205.08184* (2022).

[72] Yuki Motomura, Akira Takeshita, Yuka Egashira, Takayuki Nishimura, Yeon-kyu Kim, and Shigeki Watanuki. 2015. Interaction between valence of empathy and familiarity: is it difficult to empathize with the positive events of a stranger? *Journal of physiological anthropology* 34 (2015), 1–9.

[73] Moin Nadeem, Anna Bethke, and Siva Reddy. 2020. StereoSet: Measuring stereotypical bias in pretrained language models. *arXiv preprint arXiv:2004.09456* (2020).

[74] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *2013 IEEE Symposium on Security and Privacy*. IEEE, 541–555.

[75] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs* 41, 1 (2007), 100–126.

[76] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.

[77] Yong Jin Park. 2013. Digital literacy and privacy behavior online. *Communication research* 40, 2 (2013), 215–236.

[78] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[79] Iryna Pentina, Lixuan Zhang, Hatem Bata, and Ying Chen. 2016. Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison. *Computers in Human Behavior* 65 (2016), 409–419.

[80] Jenny Preece. 1999. Empathic communities: Balancing emotional and factual communication. *Interacting with computers* 12, 1 (1999), 63–77.

[81] Stephanie D Preston and Frans BM De Waal. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences* 25, 1 (2002), 1–20.

[82] Dixon Prem Daniel Rajendran and Rangaraja P Sundarraj. 2021. Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings. *International Journal of Information Management Data Insights* 1, 2 (2021), 100027.

[83] Robert W Reeder, Iulia Ion, and Sunny Consolvo. 2017. 152 simple steps to stay safe online: Security advice for non-tech-savvy users. *IEEE Security & Privacy* 15, 5 (2017), 55–64.

[84] Paul Resnick and Hal R Varian. 1997. Recommender systems. *Commun. ACM* 40, 3 (1997), 56–58.

[85] Frantz Rowe. 2020. Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world. *International Journal of Information Management* 55 (2020), 102178.

[86] Christoph Schneider, Markus Weinmann, and Jan vom Brocke. 2018. Digital Nudging: Guiding Online User Choices through Interface Design. *Commun. ACM* 61, 7 (jun 2018), 67–73. https://doi.org/10.1145/3213765

[87] Donghee Shin. 2018. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in human behavior* 78 (2018), 64–73.

[88] Maria Sideri, Angeliki Kitsiou, Eleni Tzortzaki, Christos Kalloniatis, and Stefanos Gritzalis. 2019. Enhancing university students' privacy literacy through an educational intervention: a Greek case-study. *International Journal of Electronic Governance* 11, 3-4 (2019), 333–360.

[89] Jellie Sierksma, Jochem Thijs, and Maykel Verkuyten. 2015. In-group bias in children's intention to help can be overpowered by inducing empathy. *British Journal of Developmental Psychology* 33, 1 (2015), 45–56.

[90] Deborah A Small and George Loewenstein. 2003. Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and uncertainty* 26 (2003), 5–16.

[91] Eric Michael Smith, Melissa Hall, Melanie Kambadur, Eleonora Presani, and Adina Williams. 2022. "I'm sorry to hear that": Finding New Biases in Language Models with a Holistic Descriptor Dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 9180–9211.

[92] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.

[93] Markus Spiekermann. 2019. Data marketplaces: Trends and monetisation of data goods. *Intereconomics* 54, 4 (2019), 208–216.

[94] Jessica Staddon, Alessandro Acquisti, and Kristen LeFevre. 2013. Self-reported social network behavior: Accuracy predictors and implications for the privacy paradox. In *2013 International Conference on Social Computing*. IEEE, 295–302.

[95] Antti Surma-Aho and Katja Hölttä-Otto. 2022. Conceptualization and operationalization of empathy in design research. *Design Studies* 78 (2022), 101075.

[96] Juliana Sutanto, Elia Palme, Chuan-Hoo Tan, and Chee Wei Phang. 2013. Addressing the personalization-privacy paradox: An empirical assessment from a field experiment on smartphone users. *MIS quarterly* (2013), 1141–1164.

[97] John K Tarus, Zhendong Niu, and Abdallah Yousif. 2017. A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems* 72 (2017), 37–48.

[98] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N Bazarova. 2019. Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[99] Helma Van Rijn, Froukje Sleeswijk Visser, Pieter Jan Stappers, and Aslı Deniz Özakar. 2011. Achieving empathy with users: the effects of different sources of information. *CoDesign* 7, 2 (2011), 65–77.

[100] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[101] Meredydd Williams, Jason RC Nurse, and Sadie Creese. 2016. The perfect storm: The privacy paradox and the Internet-of-Things. In *2016 11th International Conference on Availability, Reliability and Security (ARES)*. IEEE, 644–652.

[102] Pamela J Wisniewski, Bart P Knijnenburg, and Heather Richter Lipford. 2017. Making privacy personal: Profiling social network users to inform privacy education and nudging. *International Journal of human-computer studies* 98 (2017), 95–108.

[103] Peter Wright and John McCarthy. 2008. Empathy and experience in HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 637–646.

[104] Philip Fei Wu. 2019. The privacy paradox in the context of online social networking: A self-identity perspective. *Journal of the Association for Information Science and Technology* 70, 3 (2019), 207–217.

[105] Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk models of online behavioral advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1957–1969.

[106] Eric Zeng, Rachel McAmis, Tadayoshi Kohno, and Franziska Roesner. 2022. What factors affect targeting and bids in online advertising? a field measurement study. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 210–229.

[107] Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. 2020. Examining the adoption and abandonment of security, privacy, and identity theft protection practices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[108] Johannes Zrenner, Frederik Oliver Möller, Christian Jung, Andreas Eitel, and Boris Otto. 2019. Usage control architecture options for data sovereignty in business ecosystems. *Journal of Enterprise Information Management* 32, 3 (2019), 477–495.

A PROMPTS AND FEW-SHOT LEARNING EXAMPLES

A.1 Prompts and few-shot learning examples for generating personas using our approach

Prompt to generate persona description

Return a realistic profile. This year is 2023. The income should be in dollars. The birthday should be in the MM/DD/YYYY format. The demographic of this person should represent the US population sample.
The generated profile should match the following guidance: <guidance>.
Fit into the braces in the profile:
{First name} {Last name} is a {age ranging from 18 to 70 subject to continuous uniform distribution} {race} {gender} living in {real home address with street, city, state, and zip code}. {Pronoun} speaks {spoken language}. Pronoun's education background is {educational background}. {Pronoun}'s date of birth is {date of birth}. {Pronoun} is a {occupation}, and the annual income is {income in dollar}. {marital status} {parental status} {detailed habits and preferences when using the computer, mobile phone, and the Internet}.

The format of the generated result should look like the following examples: <few-shot learning example>
Return the profile in only one paragraph.

Few-shot learning example for the generation of persona description

Abigail Patel is a 32-year-old Asian American female living at 325 Main St, Newark, NJ 07102. She speaks English and her educational background includes a bachelor's degree in Marketing. Abigail's date of birth is 05/26/1991. She is currently working as a marketing manager, with an annual income of $85,000. Abigail is married and has two children. She enjoys browsing social media and streaming movies on her mobile phone during her free time. When using her computer, she prefers using a wireless mouse and keyboard for easy navigation. On the internet, she likes to shop for clothes and read reviews before making a purchase.

Prompt to generate privacy attributes

<few-shot learning example>
Given the profile: <persona>.
Return the attributes in this format:
{“first name”: “”, “last name”: “”, “age”: “”, “gender”: “”, “race”: “”, “street”: “”, “city”: “”, “state”: “”, "zip code": “”, “spoken language”: “”, “educational background”: “”, “birthday”: “”, “job”: “”, “income”: “”, “marital status”: “”, “parental status”: “”, “online behavior”: “”}

Few-shot learning example for the generation of privacy attributes

Given the profile: Abigail Patel is a 32-year-old Asian American female living at 325 Main St, Newark, NJ 07102. She speaks English and her educational background includes a bachelor’s degree in Marketing. Abigail’s date of birth is 05/26/1991. She is currently working as a marketing manager, with an annual income of $85,000. Abigail is married and has two children. She enjoys browsing social media and streaming movies on her mobile phone during her free time. When using her computer, she prefers using a wireless mouse and keyboard for easy navigation. On the internet, she likes to shop for clothes and read reviews before making a purchase.
Return the attributes in this format:
{“first name”: “Abigail”, “last name”: “Patel”, “age”: “32”, “gender”: “female”, “race”: “Asian American”, “street”: “325 Main St”, “city”: “Newark”, “state”: “NJ”, "zip code": “07102”, “spoken language”: “English”, “educational background”: “bachelor’s degree in Marketing”, “birthday”: “05/26/1991”, “job”: “marketing manager”, “income”: “85,000”, “marital status”: “married”, “parental status”: “has two children”, “online behavior”: “She enjoys browsing social media and streaming movies on her mobile phone during her free time. When using her computer, she prefers using a wireless mouse and keyboard for easy navigation. On the internet, she likes to shop for clothes and read reviews before making a purchase.”}

Prompt for generating portrait image prompt

Given the profile description, output a descriptive prompt to generate a realistic human head portrait image, limit the word to 30 words: {persona description}

Prompt to generate device and browser

Given the profile: {persona}, infer the browser and device the person uses:

Prompt to generate schedule

<few-shot learning example>
You are acting as a game event designer. Write daily events for this person: {persona description}. Show me a reasonable schedule for this person from {start_date} to {end_date}. The life in the period is similar to 2021. You can generate fake but reasonable data that is related to the profile. The start time of one day is 00:00:00. Generate events from 00:00:00 to 23:59:59 for each day.
Return a list of dict.
Output the following JSON format in plain text:
{ “start time”: <start moment of the event>, “end time”: <start moment of the event>, “event”: <event> }
Never provide additional context.

Few-shot learning example for the generation of schedule

"profile": "Daniel Chan is a 30yearold Asian man living in Seattle, Washington with zip code 98101. He is a project manager, and the annual income is around one hundred and twenty thousand USD. He lives alone in a studio apartment and likes to keep his space clean and organized. In his free time, he enjoys playing video games and reading books on his Kindle. He also likes to use social media platforms such as Twitter and Reddit to keep up with the latest news and trends. "
"location_history": [["2023-06-05 00:00:00", "2023-06-05 07:00:00", "Home - 1420 5th Ave, Seattle, WA 98101"], ["2023-06-05 07:00:00", "2023-06-05 08:30:00", "Golds Gym - 1220 Howell St, Seattle, WA 98101"], ["2023-06-05 08:30:00", "2023-06-05 09:00:00", "Starbucks - 1125 4th Ave, Seattle, WA 98101"]]

Prompt to generate browsing history

<few-shot learning example>
Given the person's profile: {persona description}, and the schedule: {schedule}, generate {number} browser history entries from {start_date} to {end_date}.
No browsing history between 00:00:00 and 07:00:00. The webpage title should reflect the content in the webpage url. The webpage be reasonable and related to the the schedule. Don't add the address of the schedule to the webpage title. The datetime should be realistic and associated with the webpage content. The datetime second should not be 0. The datetime should be dispensed in one day.
You can generate fake but reasonable data that is consistent with the profile and schedule. Output following list format in plain text:[[<datetime>, <webpage titile>, <webpage url>],]
Never provide additional context.

Few-shot learning example for the generation of browser history

“profile”: “John Smith is a 25yearold Caucasian male living at 123 Park Ave, New York, NY 10001. He speaks English and his educational background includes studying Computer Science and Data Analysis. John’s date of birth is 09/15/1998. He is currently a student, and his annual income is $5000. John is single and does not have any children. He enjoys coding and exploring new technologies on his computer. On his mobile phone, he prefers using apps for productivity and staying up to date with the latest tech news. When using the Internet, he enjoys participating in online coding forums and watching tutorial videos to enhance his skills.”

“schedule”:[[‘2023-07-10 00:00:00’, ‘2023-07-10 07:00:00’, ‘Home - 123 Park Ave, New York], [‘2023-07-10 07:00:00’, ‘2023-07-10 08:00:00’, ‘Morning Exercise - 987 8th Ave, New York, NY 10019’], [‘2023-07-10 08:00:00’, ‘2023-07-10 09:00:00’, ‘Breakfast - 654 Hudson St, New York, NY 10014’], [‘2023-07-10 09:00:00’, ‘2023-07-10 12:00:00’, ‘Study Computer Science - 101 Lafayette St, New York, NY 10013’], [‘2023-07-10 12:00:00’, ‘2023-07-10 13:00:00’, ‘Lunch - 246 Spring St, New York, NY 10013’], [‘2023-07-10 13:00:00’, ‘2023-07-10 15:00:00’, ‘Online Coding Forums - 876 4th Ave, New York, NY 10018’], [‘2023-07-10 15:00:00’, ‘2023-07-10 17:00:00’, ‘Study Data Analysis - 321 Canal St, New York, NY 10013’], [‘2023-07-10 17:00:00’, ‘2023-07-10 18:00:00’, ‘Break - 789 6th Ave, New York, NY 10001’], [‘2023-07-10 18:00:00’, ‘2023-07-10 20:00:00’, ‘Dinner - 897 Broadway, New York, NY 10003’], [‘2023-07-10 20:00:00’, ‘2023-07-10 23:59:59’, ‘Free Time - 456 Broadway, New York, NY 10013’]],

"browser_history":[[‘2023-07-10 15:27:08’, ‘Learning Log: Consider how data analysts approach tasks’, ‘https://www.coursera.org/learn/foundations-data/supplement/I086K/learning-log-consider-how-data-analysts-approach-tasks’], [‘2023-07-10 15:24:41’, ‘Case Study: New data perspectives’, ‘https://www.coursera.org/learn/foundations-data/supplement/nhC19/case-study-new-data-perspectives’], [‘2023-07-10 15:21:14’, ‘Data analytics in everyday life’, ‘https://www.coursera.org/learn/foundations-data/lecture/N5lvQ/data-analytics-in-everyday-life’]]

Prompt to generate social media post content

Provide ideas for this person to write posts (limit the word to 140 words) based on the profile and location history: {profile} {schedule}
The schedule is in the format of [[start time, end time, address]].
Show me only num reasonable description in total between start_date and end_date to provide ideas. The life in the given time period is similar to 2021 so you can generate the description based on your current data.
You should only return the list to me without any explanation message. You don’t need to use any real-time data, just generate reasonable and consistent data. You don’t need to generate descriptions that may be inappropriate, irrelevant, or offensive. You do not need to manipulate the data in a way that is specific to a given time period. The seconds in the time should not be 00, it should be the format like 15:23:12.
Output the following JSON format in plain text: [{ "time": <time in string format>, "address": <address where this person share the life>, "content": <content>, }]
Never provide additional context.

Few-shot learning example for the generation of social media post content

“profile”: “Emily Rodriguez is a 46yearold Hispanic female living in 602 S Fairfax Ave, Los Angeles, CA 90036. She works as a nurse and earns an annual income of $70,000. Emily is happily married with two children who are currently in college. In her free time, she enjoys reading and gardening. Emily prefers using her mobile phone for browsing social media and checking emails while using her laptop for work-related tasks. She is mindful of her online security and regularly updates her passwords and privacy settings.”

“schedule”:[[‘2023-06-06 00:00:00’, ‘2023-06-06 07:30:00’, ‘Home - 123 Main St, Los Angeles, CA 90022’], [‘2023-06-06 07:30:00’, ‘2023-06-06 08:15:00’, ‘Starbucks - 5353 E Olympic Blvd, Los Angeles, CA 90022’], [‘2023-06-06 08:15:00’, ‘2023-06-06 12:00:00’, ‘Tech Office - 3000 E 1st St, Los Angeles, CA 90063’], [‘2023-06-06 12:00:00’, ‘2023-06-06 12:45:00’, ‘Lunch Spot - 3000 E 1st St, Los Angeles, CA 90063’], [‘2023-06-06 12:45:00’, ‘2023-06-06 17:30:00’, ‘Tech Office - 3000 E 1st St, Los Angeles, CA 90063’], [‘2023-06-06 17:30:00’, ‘2023-06-06 19:00:00’, ‘Gym - 1234 Whittier Blvd, Los Angeles, CA 90022’], [‘2023-06-06 19:00:00’, ‘2023-06-06 19:45:00’, ‘Grocery Store - 5432 Whittier Blvd, Los Angeles, CA 90022’], [‘2023-06-06 19:45:00’, ‘2023-06-06 21:00:00’, ‘Home - 123 Main St, Los Angeles, CA 90022’], [‘2023-06-06 21:00:00’, ‘2023-06-06 22:30:00", ‘Favorite Local Park - 5432 E 4th St, Los Angeles, CA 90022’], [‘2023-06-06 22:30:00’, ‘2023-06-06 23:59:59", ‘Home - 123 Main St, Los Angeles, CA 90022’]],

"posts": [[‘2023-06-06 07:31:42’, ‘Starting my day with a refreshing cup of coffee at Starbucks. Ready to tackle another day at work! #CoffeeLover #WorkLifeBalance’, ‘Starbucks - 5353 E Olympic Blvd, Los Angeles, CA 90022’], [‘2023-06-06 19:01:02’, ‘Feeling the burn at the gym! Taking care of my health and fitness is a top priority. #FitnessJourney #HealthyLiving’, ‘Gym - 1234 Whittier Blvd, Los Angeles, CA 90022’]]

Prompt for the prompt to generate social media post image

Given the post {content}, output a descriptive prompt to generate a realistic life image, limit the word to 30 words:

A.2 Prompts for generating personas using baseline GPT

Prompt to generate persona description

Return a realistic profile. This year is 2023. The income should be in dollars. The birthday should be in the MM/DD/YYYY format. The demographic of this person should represent the US population sample.
The generated profile should match the following guidance: <guidance>.
Fit into the braces in the profile:
{First name} {Last name} is a {age ranging from 18 to 70 subject to continuous uniform distribution} {race} {gender} living in {real home address with street, city, state, and zip code}. {Pronoun} speaks {spoken language}. Pronoun’s education background is {educational background}. {Pronoun}’s date of birth is {date of birth}. {Pronoun} is a {occupation}, and the annual income is {income in dollar}. {marital status} {parental status} {detailed habits and preferences when using the computer, mobile phone, and the Internet}.

Return the profile in only one paragraph.

Prompt to generate privacy attributes

Given the profile: <persona>.
Return the attributes in this format:
{“first name”: “”, “last name”: “”, “age”: “”, “gender”: “”, “race”: “”, “street”: “”, “city”: “”, “state”: “”, "zip code”: “”, “spoken language”: “”, “educational background”: “”, “birthday”: “”, “job”: “”, “income”: “”, “marital status”: “”, “parental status”: “”, “online behavior”: “”}

Prompt for generating portrait image

Generate a realistic human head portrait image

Prompt to generate device and browser

Generate the browser and device a person uses:

Prompt to generate schedule

You are acting as a game event designer. Write daily events for a persona. Show me a reasonable schedule for this person from {start_date} to {end_date}. The life in the period is similar to 2021. You can generate fake but reasonable data that is related to the profile. The start time of one day is 00:00:00. Generate events from 00:00:00 to 23:59:59 for each day.
Return a list of dict.
Output the following JSON format in plain text:
{ “start time”: <start moment of the event>, “end time”: <start moment of the event>, “event”: <event> }
Never provide additional context.

Prompt to generate browsing history

Generate {number} browser history entries from {start_date} to {end_date}.
No browsing history between 00:00:00 and 07:00:00. The webpage title should reflect the content in the webpage url. The webpage be reasonable and related to the the schedule. Don’t add the address of the schedule to the webpage title. The datetime should be realistic and associated with the webpage content. The datetime second should not be 0. The datetime should be dispensed in one day.
You can generate fake but reasonable data that is consistent with the profile and schedule. Output following list format in plain text:[[<datetime>, <webpage titile>, <webpage url>],]
Never provide additional context.

Prompt to generate social media post content

Provide ideas for a person to write posts (limit the word to 140 words)
Show me only num reasonable description in total between start_date and end_date to provide ideas. The life in the given time period is similar to 2021 so you can generate the description based on your current data.
You should only return the list to me without any explanation message. You don’t need to use any real-time data, just generate reasonable and consistent data. You don’t need to generate descriptions that may be inappropriate, irrelevant, or offensive. You do not need to manipulate the data in a way that is specific to a given time period. The seconds in the time should not be 00, it should be the format like 15:23:12.
Output the following JSON format in plain text: [{ "time": <time in string format>, "address": <address where this person share the life>, "content": <content>, }]
Never provide additional context.

Prompt to generate social media post image

Generate a realistic life image for social media posts

B PARTICIPANTS' DEMOGRAPHIC DATA

Persona ID	Age	Gender	Education level	Self Rated Digital Literacy
1	26	Female	Master's	5
2	26	Male	Ph.D.	5
3	25	Female	Bachelor's	5
4	25	Male	Bachelor's	3
5	28	Female	Master's	5
6	25	Female	Master's	4
7	24	Female	Master's	5
8	25	Female	Master's	4
9	20	Male	High school Diploma	5
10	33	Male	Ph.D.	5
11	26	Male	Master's	5
12	20	Male	High school Diploma	2
13	19	Female	High school Diploma	4
14	23	Female	Master's	4
15	28	Female	Master's	4

Table 4. The demographic information of participants

C CODE SYSTEM 1

This is the code system according to which the qualitative data from part one of the interview has been coded.

C.1 Codes

(1) Emotional responses
  (a) Ads
    (i) Feel confused when failing to understand the ad
    (ii) Feel excited that the ad targeting is accurate to the persona's information
  (b) personal information
    (i) Feel bored with uninteresting photos in the posts
    (ii) Feel suspicious toward overly consistent posts and events
    (iii) Feel confused about who takes the photos in the posts
    (iv) Feel satisfied & surprised when seeing diverse browsing history
    (v) Feel confused about distances between different places
    (vi) Have a higher tolerance for persona in unfamiliar fields
    (vii) Feel excited when observing consistent information
(2) Type of Non-authenticity
  (a) Inconsistent with personal experience
    (i) Personal description
      (A) income is too low
    (ii) Social media is not real
      (A) Revealing personal privacy
      (B) Sharing only factual content without emotions and thoughts

    - (C) Repetitive posts across platforms
    - (D) Too frequent and repetitive posting
  - (iii) Schedule and Time Management
    - (A) Over-organized schedule
    - (B) Unreasonable time allocations (e.g., too short work time, too long exercise time)
    - (C) Schedule being too tight
    - (D) Inconsistent work intervals
    - (E) Too few occasional events
  - (iv) Lifestyle
    - (A) Excessive or insufficient grocery shopping
    - (B) Too few occasional incidents or events
    - (C) Inconsistent income with the lifestyle presented
    - (D) Excessive exercise or lack thereof
    - (E) Inconsistent times for activities (e.g., too early gym sessions)
- (b) Consistency of Information
  - (i) Picture does not match personal info
  - (ii) Interest is inconsistent with education
  - (iii) Background posts are inconsistent with events
  - (iv) Event is inconsistent with personal description
  - (v) Inconsistent title for same link
  - (vi) Events are not consistent with profile
  - (vii) Browsing history is inconsistent with events
  - (viii) Browsing history should be more consistent with hobbies
  - (ix) URL is inconsistent with title picture does not match income
  - (x) Inconsistent events
  - (xi) Income does not match with job
  - (xii) Inconsistent location
  - (xiii) Picture does not match with content
  - (xiv) Inconsistency between picture
  - (xv) Browsing history
    - (A) Browsing history appearing at unlikely times
    - (B) Repetitive browsing behavior
    - (C) Identical timestamps in browsing history
    - (D) Browsing sites too basic for an experienced person
    - (E) No connection between browsing history and personal info
    - (F) Incorrect link title
    - (G) Lack of record in certain time period
    - (H) Ssome events should not have browsing history every day
  - (xvi) Out of context issues
    - (A) Content lacking connection to personal info
    - (B) Missing specificity and details in content

- (3) Privacy attributes contributing to authenticity
  - (a) Consistency
    - (i) Events are consistent with the job
    - (ii) With personal experience
    - (iii) Events are consistent with hobbies
    - (iv) Social media posts are consistent with weekly schedule
    - (v) Browsing history is consistent with hobbies
    - (vi) Events among a week are consistent
- (4) Strategies to modify information
  - (a) Be more specific
  - (b) Increase/decrease certain activities
  - (c) Modify work and leisure schedules
  - (d) Browsing in a progressive way
  - (e) Some events can be more dispersed
  - (f) Increase salary
  - (g) Express more diverse emotion and attitude
  - (h) More life-oriented browsing history
- (5) Reasons for irrelevant ads
  - (a) Not mentioned in the persona's data
  - (b) Inconsistent with persona's data
    - (i) Profile photos
    - (ii) Race
    - (iii) Social media post
    - (iv) Location
    - (v) Marital status
    - (vi) Job
    - (vii) Interest or hobbies
    - (viii) Income
    - (ix) Age and state
    - (x) daily activities
    - (xi) Personal property
    - (xii) Gender
  - (c) Looks like spam
  - (d) Already have or know similar things
  - (e) Ad feels too generic
  - (f) Choose specific websites instead of ad
- (6) Perceived privacy attributes affecting ads
  - (a) Gender
  - (b) Social media post
  - (c) Personal property
  - (d) Income

(e) Parental status
(f) Job related location
(g) Daily activities
(h) Marital status
(i) Interests and hobbies
(j) Race
(k) Age and state

D PERSONAS GENERATED BY OUR APPROACH

John Smith is an **18-year-old White male** living in 123 Oak St, **San Francisco**, CA 94102. He speaks English and his education background is a high school diploma. John's date of birth is 09/15/2005. He is currently working as a **cashier**, with an annual income of $20,000. John is single and does not have any children. He enjoys playing video games and watching YouTube videos on his computer. On his mobile phone, he spends most of his time scrolling through social media and texting his friends. When using the internet, he likes to research the latest technology gadgets and watch streaming services.

Carlos Rodriguez is a **30-year-old Hispanic male** living in 456 Elm Street, **Los Angeles**, CA 90005. He speaks both English and Spanish fluently. Carlos's education background includes a bachelor's degree in Business Administration. Carlos's date of birth is 09/14/1993. He is currently working as a **financial analyst**, with an annual income of $75,000. Carlos is single and does not have any children. He enjoys using his computer for online gaming and watching videos on YouTube. On his mobile phone, Carlos frequently uses social media to connect with friends and family. When browsing the Internet, he enjoys reading news articles and checking sports scores.

Michael Johnson is a **42-year-old African American male** living in 123 Oak Street, **Atlanta**, GA 30303. He speaks English and his education background includes a bachelor's degree in Business Administration. He is currently working as a **financial analyst**, with an annual income of $70,000. Michael is single and does not have any children. He enjoys using his computer for online gaming and browsing news websites. When using his mobile phone, he prefers to use social media and watch videos. On the internet, he likes to research investment opportunities and read financial blogs.

Andre Johnson is a **51-year-old African American male** living in 456 Oak St, **Birmingham**, AL 35203. He speaks English and his education background includes a high school diploma. Andre's date of birth is 08/12/1972. He is currently working as a **custodian**, with an annual income of $35,000. Andre is married and has one child. He enjoys using his computer for basic tasks such as checking emails and browsing news websites. On his mobile phone, he mainly uses it for making calls and sending text messages. When using the internet, he prefers to use social media to stay connected with friends and family.

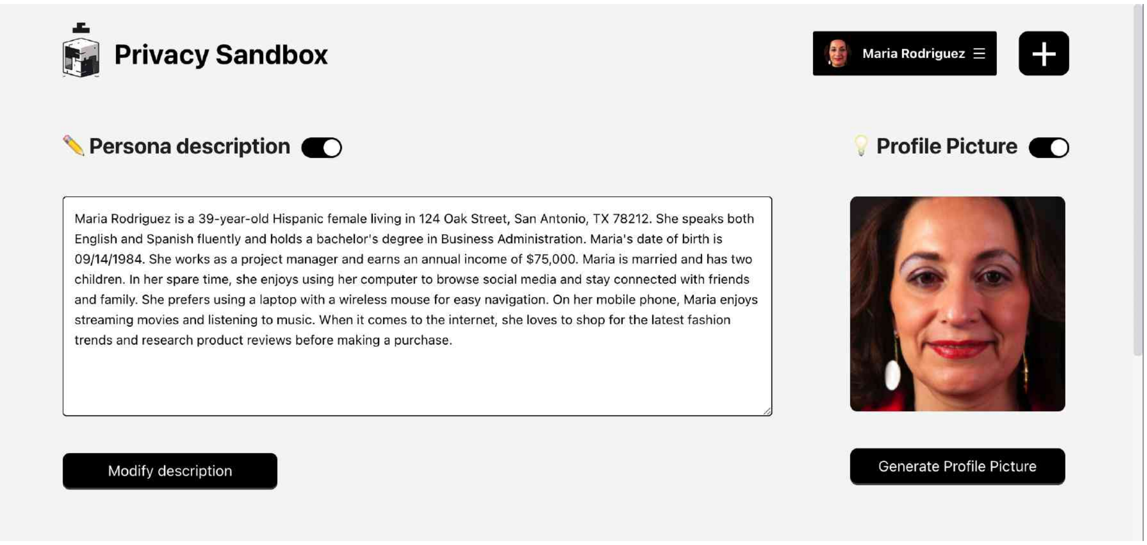
Emily Wilson is a **22-year-old White female** living in 215 Elm St, **San Francisco**, CA 94102. She speaks English and her education background includes a bachelor's degree in Psychology. Emily's date of birth is 09/14/2001. She is currently working as a **research assistant**, with an annual income of $50,000. Emily is single and has no children. She enjoys using her computer for graphic design projects and uses a gaming mouse for precise movements. On her mobile phone, she uses social media to stay connected with friends and family. While browsing the Internet, she likes to read articles on psychology topics and participate in online forums for discussions.

Isabella Johnson is a **36-year-old White female** living in 123 Maple St, **Seattle**, WA 98101. She speaks English and her education background includes a high school diploma. Isabella's date of birth is 07/14/1987. She is currently working as a **retail sales associate**, with an annual income of $30,000. Isabella is single and has no children. She enjoys browsing social media and watching videos on her mobile phone during her free time. When using her computer, she prefers using a wireless mouse and keyboard for easy navigation. On the internet, she likes to read news articles and participate in online forums.

Maria Rodriguez is a **39-year-old Hispanic female** living in 124 Oak Street, **San Antonio**, TX 78212. She speaks both English and Spanish fluently and holds a bachelor's degree in Business Administration. Maria's date of birth is 09/14/1984. She works as a **project manager** and earns an annual income of $75,000. Maria is married and has two children. In her spare time, she enjoys using her computer to browse social media and stay connected with friends and family. She prefers using a laptop with a wireless mouse for easy navigation. On her mobile phone, Maria enjoys streaming movies and listening to music. When it comes to the internet, she loves to shop for the latest fashion trends and research product reviews before making a purchase.

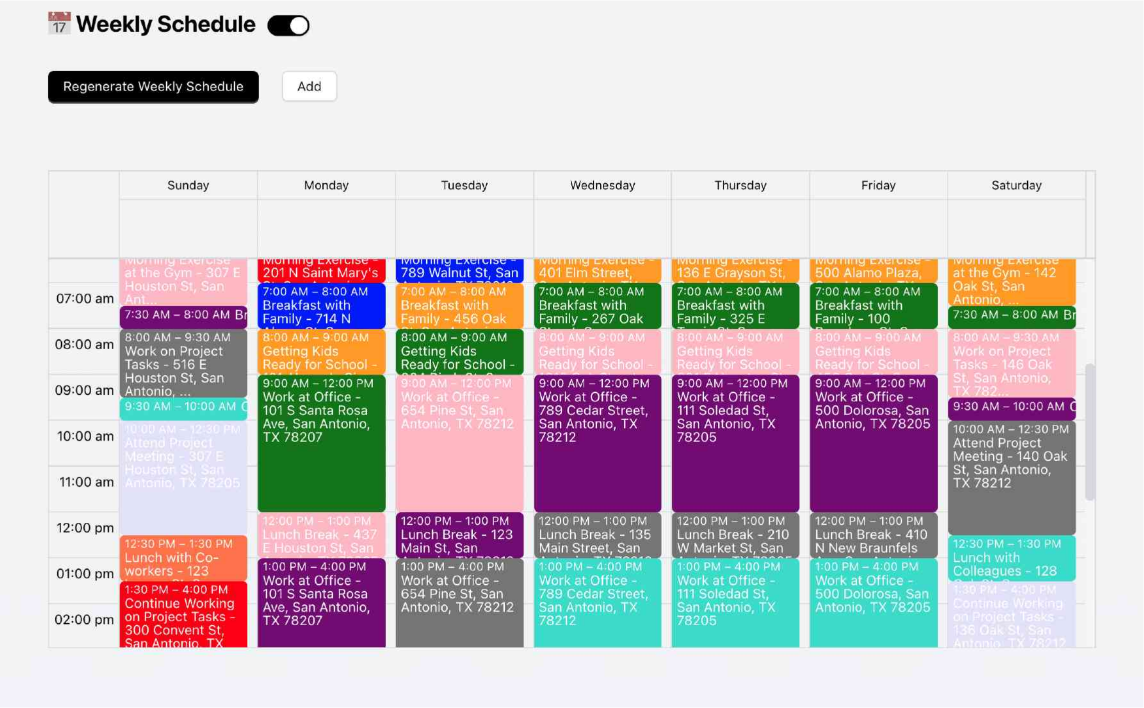
Linda Nguyen is a **53-year-old Asian American female** living in 26 Oak St, **San Francisco**, CA 94102. She speaks English and her education background includes a bachelor's degree in Fine Arts. She is currently working as a senior **graphic designer**, with an annual income of $125,000. Linda is married and has two children. She enjoys designing and editing graphics on her computer, using professional software and a stylus for precise control. On her mobile phone, she likes to browse social media and play puzzle games. When using the internet, she enjoys researching design trends and reading articles about art and creativity.

Fig. 8. Personas generated by our approach.



Maria Rodriguez is a 39-year-old Hispanic female living in 124 Oak Street, San Antonio, TX 78212. She speaks both English and Spanish fluently and holds a bachelor's degree in Business Administration. Maria's date of birth is 09/14/1984. She works as a project manager and earns an annual income of $75,000. Maria is married and has two children. In her spare time, she enjoys using her computer to browse social media and stay connected with friends and family. She prefers using a laptop with a wireless mouse for easy navigation. On her mobile phone, Maria enjoys streaming movies and listening to music. When it comes to the internet, she loves to shop for the latest fashion trends and research product reviews before making a purchase.

Fig. 9. Screenshot of persona description of Maria



Sunday

- Morning Exercise at the Gym - 307 E Houston St, San Ant...
- 7:30 AM – 8:00 AM Br
- 8:00 AM – 9:30 AM Work on Project Tasks - 516 E Houston St, San Antonio, ...
- 9:30 AM – 10:00 AM C
- 10:00 AM – 12:30 PM Attend Project Meeting - 307 E Houston St, San Antonio, TX 78205
- 12:30 PM – 1:30 PM Lunch with Co-workers - 123
- 1:30 PM – 4:00 PM Continue Working on Project Tasks - 300 Convent St, San Antonio, TX

Monday

- Morning Exercise - 201 N Saint Mary's
- 7:00 AM – 8:00 AM Breakfast with Family - 714 N
- 8:00 AM – 9:00 AM Getting Kids Ready for School -
- 9:00 AM – 12:00 PM Work at Office - 101 S Santa Rosa Ave, San Antonio, TX 78207
- 12:00 PM – 1:00 PM Lunch Break - 437 E Houston St, San
- 1:00 PM – 4:00 PM Work at Office - 101 S Santa Rosa Ave, San Antonio, TX 78207

Tuesday

- Morning Exercise - 789 Walnut St, San
- 7:00 AM – 8:00 AM Breakfast with Family - 456 Oak
- 8:00 AM – 9:00 AM Getting Kids Ready for School -
- 9:00 AM – 12:00 PM Work at Office - 654 Pine St, San Antonio, TX 78212
- 12:00 PM – 1:00 PM Lunch Break - 123 Main St, San
- 1:00 PM – 4:00 PM Work at Office - 654 Pine St, San Antonio, TX 78212

Wednesday

- Morning Exercise - 401 Elm Street,
- 7:00 AM – 8:00 AM Breakfast with Family - 267 Oak
- 8:00 AM – 9:00 AM Getting Kids Ready for School -
- 9:00 AM – 12:00 PM Work at Office - 789 Cedar Street, San Antonio, TX 78212
- 12:00 PM – 1:00 PM Lunch Break - 135 Main Street, San
- 1:00 PM – 4:00 PM Work at Office - 789 Cedar Street, San Antonio, TX 78212

Thursday

- Morning Exercise - 136 E Grayson St,
- 7:00 AM – 8:00 AM Breakfast with Family - 325 E
- 8:00 AM – 9:00 AM Getting Kids Ready for School -
- 9:00 AM – 12:00 PM Work at Office - 111 Soledad St, San Antonio, TX 78205
- 12:00 PM – 1:00 PM Lunch Break - 210 W Market St, San
- 1:00 PM – 4:00 PM Work at Office - 111 Soledad St, San Antonio, TX 78205

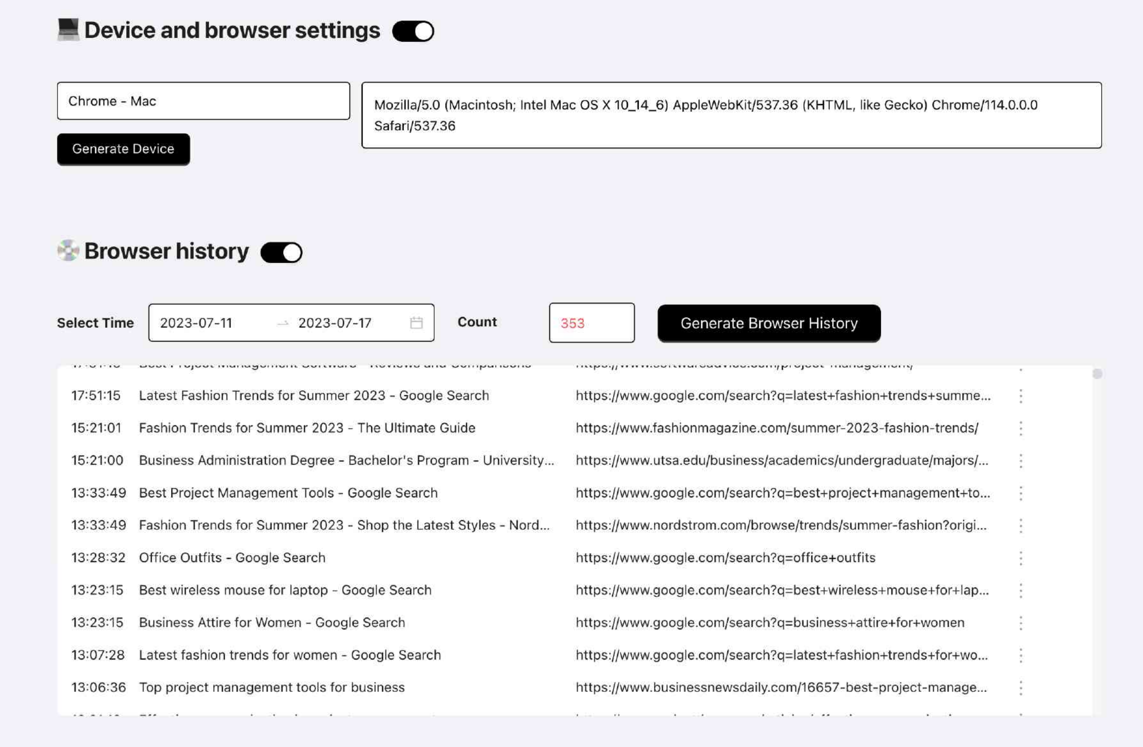
Friday

- Morning Exercise - 500 Alamo Plaza,
- 7:00 AM – 8:00 AM Breakfast with Family - 100
- 8:00 AM – 9:00 AM Getting Kids Ready for School -
- 9:00 AM – 12:00 PM Work at Office - 500 Dolorosa, San Antonio, TX 78205
- 12:00 PM – 1:00 PM Lunch Break - 410 N New Braunfels
- 1:00 PM – 4:00 PM Work at Office - 500 Dolorosa, San Antonio, TX 78205

Saturday

- Morning Exercise at the Gym - 142 Oak St, San Antonio, ...
- 7:30 AM – 8:00 AM Br
- 8:00 AM – 9:30 AM Work on Project Tasks - 146 Oak St, San Antonio, TX 782...
- 9:30 AM – 10:00 AM C
- 10:00 AM – 12:30 PM Attend Project Meeting - 140 Oak St, San Antonio, TX 78212
- 12:30 PM – 1:30 PM Lunch with Colleagues - 128
- 1:30 PM – 4:00 PM Continue Working on Project Tasks - 136 Oak St, San Antonio, TX 78212

Fig. 10. Screenshot of schedule of Maria



Chrome - Mac

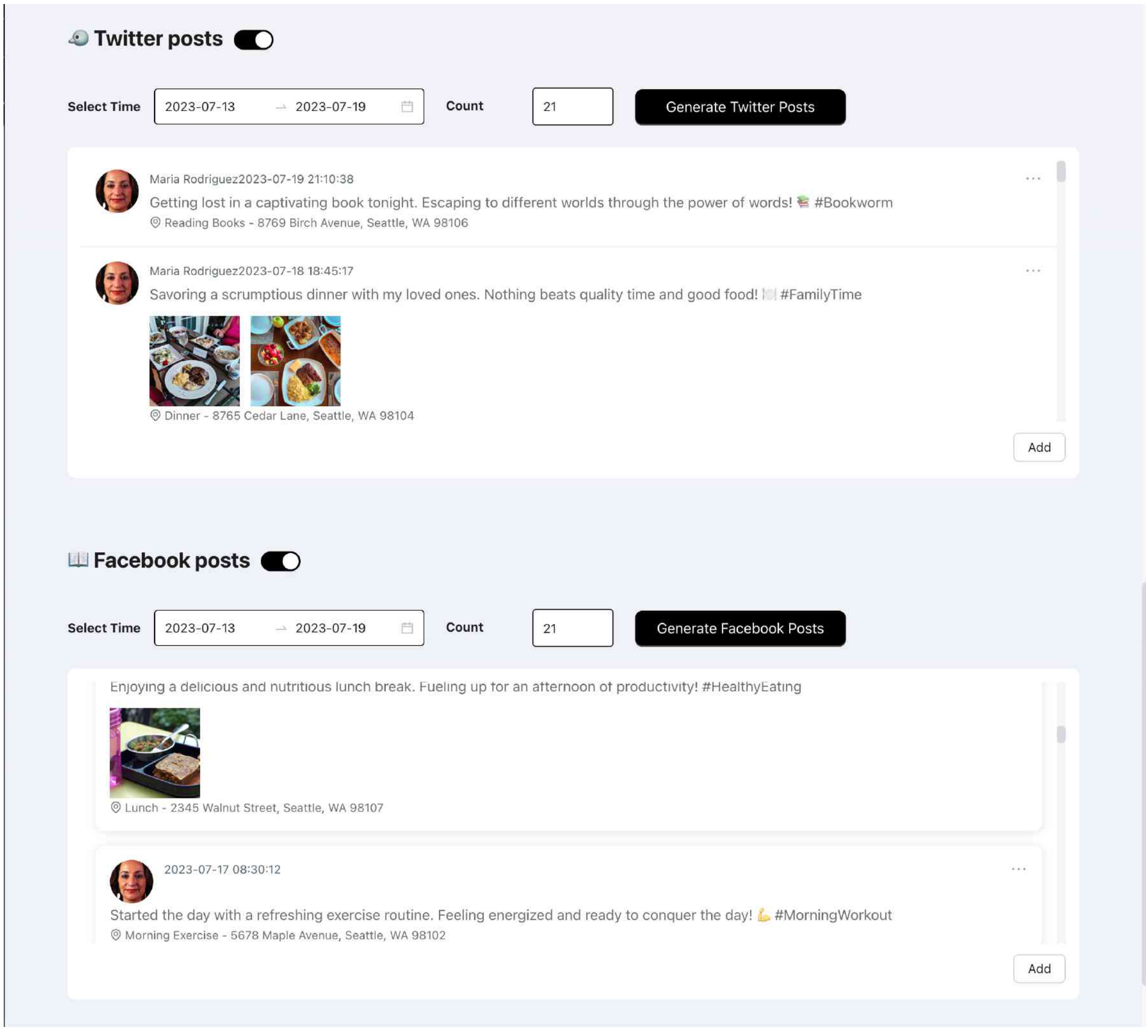
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/114.0.0.0 Safari/537.36

Select Time 2023-07-11 → 2023-07-17

Count 353

17:51:15	Latest Fashion Trends for Summer 2023 - Google Search	https://www.google.com/search?q=latest+fashion+trends+summe...
15:21:01	Fashion Trends for Summer 2023 - The Ultimate Guide	https://www.fashionmagazine.com/summer-2023-fashion-trends/
15:21:00	Business Administration Degree - Bachelor's Program - University...	https://www.utsa.edu/business/academics/undergraduate/majors/...
13:33:49	Best Project Management Tools - Google Search	https://www.google.com/search?q=best+project+management+to...
13:33:49	Fashion Trends for Summer 2023 - Shop the Latest Styles - Nord...	https://www.nordstrom.com/browse/trends/summer-fashion?origi...
13:28:32	Office Outfits - Google Search	https://www.google.com/search?q=office+outfits
13:23:15	Best wireless mouse for laptop - Google Search	https://www.google.com/search?q=best+wireless+mouse+for+lap...
13:23:15	Business Attire for Women - Google Search	https://www.google.com/search?q=business+attire+for+women
13:07:28	Latest fashion trends for women - Google Search	https://www.google.com/search?q=latest+fashion+trends+for+wo...
13:06:36	Top project management tools for business	https://www.businessnewsdaily.com/16657-best-project-manage...

Fig. 11. Screenshot of device and browsing history of Maria



🐦 Twitter posts

Select Time 2023-07-13 → 2023-07-19

Count 21

Maria Rodriguez2023-07-19 21:10:38

Getting lost in a captivating book tonight. Escaping to different worlds through the power of words! 📚 #Bookworm

Reading Books - 8769 Birch Avenue, Seattle, WA 98106

Maria Rodriguez2023-07-18 18:45:17

Savoring a scrumptious dinner with my loved ones. Nothing beats quality time and good food! 🍽️ #FamilyTime

Dinner - 8765 Cedar Lane, Seattle, WA 98104

📖 Facebook posts

Select Time 2023-07-13 → 2023-07-19

Count 21

Enjoying a delicious and nutritious lunch break. Fueling up for an afternoon of productivity! #HealthyEating

Lunch - 2345 Walnut Street, Seattle, WA 98107

2023-07-17 08:30:12

Started the day with a refreshing exercise routine. Feeling energized and ready to conquer the day! 💪 #MorningWorkout

Morning Exercise - 5678 Maple Avenue, Seattle, WA 98102

Fig. 12. Screenshot of social posts of Maria