BIS Working Papers

No 1333

Generative AI for surveys on payment apps: AI views on privacy and technology

by Koji Takahashi and Joon Suk Park

Monetary and Economic Department

March 2026

JEL classification: M31, C83, C45, D12, L86

Keywords: ChatGPT, generative artificial agents, privacy paradox, Westin index, survey, payments

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the BIS or its member central banks.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Generative AI for Surveys on Payment Apps: AI Views on Privacy and Technology [*]

Koji Takahashi[†] Joon Suk Park [‡]

March 2, 2026

Abstract

This study uses ChatGPT to simulate survey responses about payment apps, focusing on privacy and perceived benefits. By designing prompts that mirror real user characteristics, the generated responses align with findings from a Dutch survey, especially when grouped by privacy concern. Privacy-concerned agents view apps less favorably, while users show more positive attitudes than non-users, even without such traits in the prompt. However, ChatGPT fails to match the real survey's response variability and tends to overstate privacy concerns. These results indicate that generative AI can complement but not replace human surveys for studying perceptions of payment tools.

Keywords: ChatGPT, generative artificial agents, privacy paradox, Westin index, survey, payments

JEL Classification codes: M31, C83, C45, D12, L86

[*]The views expressed in this paper are those of the authors and do not necessarily reflect the official views of Bank for International Settlements, the Bank of Japan or the Bank of Korea. The authors thank Nicole Jonker and Hans Brits for kindly providing data, and Jon Frost, Hyun Song Shin, Anneke Kosse, Douglas Araujo, Fernando Perez-Cruz, Daisuke Ikeda, Shingo Watanabe, Dongjae Jung and participants at the seminars at the Bank for International Settlements and the Bank of Japan for helpful comments. All remaining errors are our own.

[†]Bank of Japan. Email: kouji.takahashi-2@boj.or.jp

[‡]Bank of Korea. Email: parkjs@bok.or.kr

1 Introduction

In this decade, we have witnessed the rapid and transformative evolution of artificial intelligence (AI). In some specific tasks, researchers report that the AI's intelligence level has reached a level comparable to that of human beings (Hendrycks et al. (2020); Choi et al. (2021); Kung et al. (2023); Cornelli et al. (2023)). Not only performing predetermined calculations, a new type of AI, known as generative AI (GenAI), is able to generate texts, pictures, music, and computer code with minimal task descriptions.

A growing number of models and applications of generative AI (GenAI) have been developed by researchers, companies, and various institutions. One potential application of GenAI is as a tool for market surveys. Market surveys require significant costs and time for implementation. In addition, it is not easy to obtain responses from samples that accurately represent the target population. Some studies show that responses by GenAI exhibit similar tendencies to those of human beings in terms of both rationality and irrationality. Moreover, previous studies, including Horton (2023), show that GenAI responds in a manner consistent with theoretical predictions, although it has some limitations in reasoning or logical thinking skills (Binz and Schulz (2023); Perez-Cruz and Shin (2024)).

In this paper, we apply GenAI as generative agents to survey the usage of payment apps, with a particular focus on perceptions of privacy and benefits. To ensure the validity of survey responses generated by GenAI, we compare them with the empirical findings of Brits and Jonker (2023), who document that Dutch consumers' use of financial apps largely reflects a rational privacy calculus—carefully weighing perceived benefits against privacy risks rather than exhibiting paradoxical behavior. This comparison allows us to evaluate whether GenAI reproduces the main behavioral patterns observed in real survey data. To the best of our knowledge, this is the first study to apply GenAI to a survey on payment technology. Privacy and data protection take center stage when a new payment technology is introduced (ECB (2023); Li (2023)). However, user perceptions of privacy issues are inherently complex. Previous studies report that people often have seemingly contradictory attitudes and opinions on privacy, a tendency known as the privacy paradox This has been extensively studied by researchers in economics, psychology, and computer science (Williams et al. (2016); Chen et al. (2022); Goldfarb and Que (2023)). Therefore, obtaining a deeper understanding of privacy perceptions is essential for policy makers to implement regulations surrounding new technologies such as payment apps.

Our study suggests that GenAI has the potential to serve as a complementary tool to conduct surveys by taking into account the complex tendencies of human beings regarding privacy issues. Specifically, GenAI could be used to help researchers brainstorm in order to create survey questions and conduct simulations of surveys before embarking on surveys with actual human beings. Using ChatGPT-4o for simulating surveys, first, we find that generative agents' views on payment app benefits and privacy are similar to actual survey results. Privacy-concerned agents view financial apps less favorably and see more risks, even without specifying this tendency in prompts. Second, the views of generative agents on the benefits of financial apps are more widely dispersed than those on risk, which is consistent with actual survey results. Third, ChatGPT provides responses that account for the differences between users and non-users of payment apps, which are also observed in the actual survey, without specifying those characteristics in the prompt. In simulations with ChatGPT, users find financial apps more beneficial while perceiving less risk than non-users. These results suggest that, to improve responses in the simulated survey with ChatGPT, it would be beneficial to include partial information about the perceptual or behavioral characteristics of generative agents in the prompt, especially when correlations between these characteristics are expected. Overall, we find evidence for the potential of GenAI to be used in market surveys. It is important to note that the training data cutoff for ChatGPT-4o predates the publication of Brits and Jonker (2023), which ensures no training leakage, as highlighted by Ludwig et al. (2025).[1]

However, there are caveats when generative agents are applied to surveys. First, the generative agents do not generate as much variation as the responses of actual humans. This limited variation fails to capture the wide range of responses found in actual surveys. Another caveat is that most generative agents are classified as "privacy fundamentalists," which is inconsistent with the actual survey. In other words, ChatGPT generates a bias in views on privacy and risks related to payment apps. To examine the source of this bias, we simulated another experiment by changing the residence country of generative agents from the Netherlands to the United States, resulting in a high share of privacy fundamentalists. This suggests that it is challenging to examine and rectify factors contributing to this bias. Therefore, users of GenAI should recognize the possibility of biases and the difficulty of eliminating them. These findings indicate that the use of GenAI for surveys requires significant caution. In addition, specifying demographics would not result in much variation in the responses, although in actual surveys, demographic information is often used for sampling. Finally, we find that specifying many different features simultaneously in prompts causes ChatGPT to put less weight on some features and more weight on others in an unexpected manner. This means that simply specifying many detailed

[1]Brits and Jonker (2023) was published in November 2023, while the training data cutoff date for ChatGPT-4o was October 2023 (OpenAI (2024)) at the time we ran simulations in December 2024. In January 2025—after we ran our simulations—OpenAI announced that the training data cutoff had been updated to June 2024 (OpenAI (2025)).

personas does not necessarily help to generate better synthetic surveys.

The rest of the paper is organized as follows. Section 2 discusses the literature. In Section 3, we illustrate a general methodology for applying GenAI to market surveys. In Section 4, we introduce the study by Brits and Jonker (2023) that we replicate and report our replication results, comparing the results from GenAI with the survey on humans. Section 5 discusses the advantages and disadvantages of using GenAI for market surveys. Finally, Section 6 provides concluding remarks.

2 Literature review

Our paper is mainly related to three strands of literature. The first strand of studies investigates economic experiments using GenAI. Using a large language model (LLM), Horton (2023) and Ma et al. (2023) implemented economic experiments, which are motivated by classic experiments in the behavioral economic literature. They find that ChatGPT can generate similar results to the original studies. Their findings suggest that ChatGPT is able to replicate human cognitive processes in a number of ways. For example, ChatGPT can make inferences from limited information, and it can learn from experience. Additionally, the results from ChatGPT are consistent with behavioral economics theory, which suggests that people are not always rational decision-makers. On the other hand, Binz and Schulz (2023) show that ChatGPT 3 fails in a causal reasoning task. In addition, Perez-Cruz and Shin (2024) show that ChatGPT 4 displays a distinctive and revealing pattern of failure in solving a logical puzzle that demands reasoning about the knowledge of others and about counterfactuals. Perez-Cruz and Shin (2024) point out that users should be careful when they use LLMs in contexts that demand reasoning in economic analysis.

Second, our paper is closely related to the study by Brand et al. (2023), which explores the possibility of using generative AI for market and social science research. Brand et al. (2023) use ChatGPT to conduct market research on consumer behavior. They find that ChatGPT's responses to survey questions (e.g., willingness-to-pay) align with economic theory and well-documented patterns of consumer behavior. Additionally, they find that ChatGPT can generate insights into consumer motivations and preferences that would not be possible with traditional survey methods. Park et al. (2024) compare survey results from humans with those from generative agents by replicating the personas of each human respondent, and finds that the generative agents respond in a manner consistent with the actual human responses. Argyle et al. (2023) demonstrate that, when appropriately conditioned on rich sociodemographic backstories, the GPT-3 language model can produce samples whose response patterns closely mirror those of diverse human subpopulations—exhibiting a property they term algorithmic fidelity, which supports the use of such models as proxies for human survey respondents in social science research. On the other hand, Bisbee et al. (2024) test whether ChatGPT can generate synthetic survey data by adopting respondent personas to reproduce results from the American National Election Study, finding that while the synthetic data closely match overall mean scores, they exhibit unnaturally low variance, and are highly sensitive to prompt phrasing and model updates, thereby raising concerns about their reliability and reproducibility. The use of LLMs to support the process of designing surveys remains a topic of active discussion. Therefore, we offer new insights on the topic by extending their study to surveys on perceptions of privacy and new technologies using GenAI.

The third strand of the literature develops computational models of social systems using GenAI. Park et al. (2023) and Ghaffarzadegan et al. (2023) have developed computational models of social systems using ChatGPT. Their models are based on the idea that human interactions can be represented as a network of relationships. ChatGPT is then used to generate text that captures the dynamics of these networks. Their models have been used to study a variety of social phenomena, such as the spread of misinformation and the emergence of social norms. Our paper is in line with the motivation of Kazinnik (2023), which uses the GenAI to simulate the synthetic survey responses to different bank run scenarios. Kazinnik (2023) finds that the trend in bank deposit withdrawals across demographic categories in the simulation with GenAI is aligned with existing empirical studies.

3 Methodology for surveys using GenAI

In this section, we first explain how we generate responses to surveys with GenAI. Then, we discuss how to match the distribution of the targeted population. In our case, the targeted distribution corresponds to the distribution of samples in Brits and Jonker (2023). Finally, we show a typical prompt in a market survey with GenAI.

3.1 Distributional information and probabilistic model

The purpose of conducting a market survey is not limited to obtaining an average or representative view on user preference for a certain feature of products or services. It can also be used to understand the multivariate aspects of users' perceptions by providing the distribution of users' preferences. Responses to a specific question could vary across people with different demographic, cultural, and psychological characteristics and in different circumstances. Furthermore, the way of answering a question can be affected by the environment surrounding them. The distributional information is useful not only for conducting a statistical test but also for understanding the implications for new products or services. In particular, from the viewpoint of policy makers, it is essential to incorporate overall impacts on users or consumers, even in the tails of distributions, since their goal is to serve all of society.

There are two main ways to generate distributions in responses by GenAI. The first is to use the probabilistic aspect of GenAI, by modifying the value of parameter called temperature that determines the randomness of responses. The second is to specify personas by introducing variation. Each of these aspects is discussed in turn.

GenAI is developed to perform creative tasks such as generating music, stories, and movies. To produce variation in their creative works, they incorporate a probabilistic aspect in their background model (Higham et al. (2023)). Without some randomness, they would always produce the same output with the same prompt, limiting their ability to create new products.

It is important to note that this feature is entirely different from a deterministic model, such as a calculator, which is designed to always provide the same answer to a specific mathematical problem. To let GenAI have flexibility in the extent of the randomness, many GenAI models allow users to adjust certain parameters. One of the parameters most related to the randomness of answers is the temperature parameter, which alters the distribution shape of possible answers in the background model.[2] The range of the temperature parameter is from 0 to 2.0. A lower value of the temperature implies a transformation of the distribution to a more skewed shape.[3]

In addition, top_p is a parameter related to the probabilistic response of GenAI. top_p determines a set of choices by setting the cutoff level of the cumulative distribution. GenAI provides an answer from possible choices of which the cumulative probability matches the top_p when they are ordered according to the probability. Therefore, it can be set from 0 to 1.[4]

In our paper, we set the temperature to one following previous studies (Brand et al. (2023)) and set top_p as one to allow the maximum flexibility of their choice.

This feature of GenAI allows users to change the probabilistic characteristics of models to fit the purpose of the usage. For example, if one uses a GenAI for a customer service and expects it to provide a consistent response to a certain type of questions from a customer, the parameters can have very low values. On the other hand, if one exploits GenAI to brainstorm a new idea

[2]For the tips for setting the parameters, see https://community.openai.com/t/cheat-sheet-mastering-temperature-and-top-p-in-chatgpt-api/172683, for example.

[3]More precisely, the temperature is a scale parameter of the softmax function that transforms the model's intermediate output into probabilities. A lower temperature value causes the GenAI model to become more deterministic, resulting in similar responses.

[4]There are other parameters such as frequency_penalty and presence_penalty that would affect the variation of responses from GenAI. We choose to examine only the effect of temperature and prompt design on the randomness of responses from GenAI. Consequently, to ensure this focus and reduce experimental noise, frequency_penalty and presence_penalty are set to 0.0.

of an advertising picture, it can be set to a high value close to one.

3.2 Variation in personas

The other way to generate variation is to change the persona set in prompts. Even though GenAI generates responses based on probabilistic models, it may not provide as much variation as responses in surveys on humans do.

For example, suppose we ask a generative agent about the usefulness of a new payment system on a Likert scale from 1 (not useful at all) to 5 (very useful), specifying the respondent's nationality and gender. Hypothetically, assume that responses are concentrated in 4 and 5. However, it is important to note that preferences for a new payment system are expected to vary substantially across different age groups. If the model provides responses assuming that it is asked to people in the average age group, which may be a younger group in the country when age is not specified in the prompts, the responses from the experiment could be biased.[5]

To implement a valid survey based on GenAI, we need to carefully specify personas for generative agents. In addition, the sampling strategy should coincide with the targeted population. Since perceptions of the benefits and risks of mobile payment apps differ among age groups, the proportion of survey respondents from various age groups could affect the overall survey results.

3.3 Prompts and personas

We adopt the role, task, and format (RTF) methodology to design the prompt as suggested by researchers (Li et al. (2023) and Motoki et al. (2023)). The role section specifies the persona given to a generative agent produced by GenAI. This section indicates that the agent would be a survey respondent with characteristics such as nationality, age, and gender. We assume the generative agents are living in a specific country and generate age and gender randomly depending on the structure of experiments, which will be described in the next section.

The task section specifies the survey questions regarding the perception of privacy and technology. The format section explains how the generative agent should answer the questions. We asked the generative agent to only provide the answers to the questions without any reasoning to parse the answers efficiently. After we carefully design the prompt, we send the prompt with different personas to the ChatGPT-4o model multiple times via the Python API to

[5]In other words, for features that are unspecified in prompts, the GenAI might provide responses based on the average values of those features in their model without varying them. This is because the responses generated by GenAI are not actual but probabilistically determined through the sampling from a marginal distribution with average values in state variables rather than from a joint distribution of multiple characteristic variables.

obtain sufficient answers.[6]

4 Replication of human-based surveys

4.1 Analysis of Jonker and Brits (2023)

Brits and Jonker (2023) conducted a survey on the usage of financial apps, focusing on the privacy paradox and privacy calculus. The privacy paradox refers to the apparent contradiction between individuals' concerns about their privacy and their willingness to disclose personal information online. On the other hand, the privacy calculus is based on a cost-benefit analysis, where rational users perceive the benefits to outweigh the risks.

The survey was conducted in November 2022 with 2,465 respondents in the Netherlands. The survey questions concerned the use of mobile apps and perceptions of privacy and data protection. In our paper, we mainly use two methodologies to measure the perceived risk and benefit of payment apps and respondents' views on privacy. First, we calculate the benefit and risk scores of respondents. Second, we categorize respondents into so-called Westin groups based on their answers.[7]

4.1.1 Privacy calculus

In the survey conducted by Brits and Jonker (2023), respondents were asked to answer the questions on a 5-point Likert scale to calculate the overall scores of benefits and risks associated with using mobile apps. The respondents included not only users of financial information apps, mobile payment apps, and activity tracking apps but also people who did not use these mobile apps.

As summary statistics, first, benefit scores were calculated as the unweighted arithmetic mean of the answers to three questions regarding benefits. The respondents were asked to answer those questions on a 5-point Likert scale, with 1 indicating "complete disagreement" and 5 indicating "complete agreement".

On the other hand, risk scores were calculated as the arithmetic average of the answers to seven questions regarding data protection and data privacy. Researchers asked respondents about the severity and likelihood of several possible inappropriate personal data use cases to calculate the risk score. A 5-point Likert scale ranging from 1 ("not at all threatening") to 5

[6] See Appendix for the details of our prompt and typical responses from ChatGPT (Figure A1).

[7] Westin groups are the classic segmentation of the public's privacy attitudes, first introduced by Alan F. Westin, which divides individuals into three categories based on their concern and willingness to trade personal data: Privacy Fundamentalists (high concern), Privacy Unconcerned (low concern), and Privacy Pragmatists (who weigh risks and benefits).

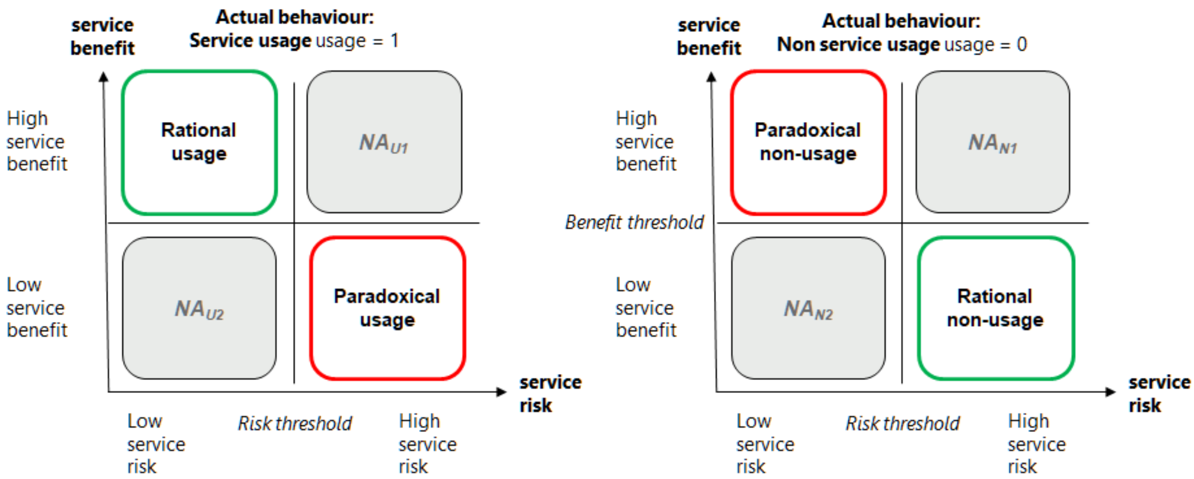

Figure 1: Illustration of rational and paradoxical user privacy behavior Gimpel et al. (2018)

Note: The figure shows the two types of behavior for user and non-user groups: rational and paradoxical. The top left of the figure indicates users who exhibit high perceived benefit and low perceived risk, and their behavior is classified as rational. In contrast, users with low perceived benefit and high perceived risk exhibit paradoxical behavior. On the right side of the figure, non-users with high perceived benefit and low perceived risk exhibit paradoxical behavior. Non-users with low perceived benefit and high perceived risk can be classified as demonstrating rational non-usage.

("extremely threatening") was used for the severity of personal data misuse. Similarly, a 5-point Likert scale ranging from 1 ("very unlikely") to 5 ("very likely") was adopted to determine the likelihood that a specific privacy breach would happen to a user or non-user of a mobile app in the next 12 months.

Brits and Jonker (2023) used the privacy paradox metric (*PPM*) developed by Gimpel et al. (2018) to calculate the proportion of Dutch consumers' use of mobile apps that could be categorized as "privacy paradoxical". The *PPM* assesses the extent of users' privacy paradox in digital services based on the theoretical construct of the privacy calculus. Figure 1 illustrates how the behavior of users and non-users is determined.

The left panel of the figure depicts the categorization of behavior of users. If a user perceives low risk and high benefit, then the user's behavior can be classified as rational usage. Conversely, if a user perceives high service risk and low service benefit, then the user's behavior becomes paradoxical usage.

The right panel of the figure demonstrates non-users' behavior. A non-user who perceives low risk and high benefit can be categorized as exhibiting paradoxical non-usage behavior. On the other hand, a non-user perceiving high risk and low benefit depicts what might be considered rational non-usage.

4.1.2 Westin type

Along with benefit and risk scores, were compared between people categorized by Westin indices, which are classifications of people's general privacy attitudes developed by Alan Westin in the late 1970s. Respondents are divided into three categories following the practice in earlier surveys (Kumaraguru and Cranor (2005)):

1. *Privacy fundamentalists* consider privacy very important and distrust organisations that ask for their personal information. These people worry about the accuracy of computerized information and additional uses made of it, and favor new laws and regulations that enhance privacy rights.

2. *Privacy unconcerned* generally trust organizations collecting their personal information and tend to forgo privacy claims to secure public-order values. These people do not favor new laws and regulations regarding privacy protection.

3. *Privacy pragmatists* seek to balance the benefits derived from consumer opportunities and public safety protections against the risks of personal information misuse. Accordingly, they demand that practical procedures are in place to ensure data accuracy, and that mechanisms exist for the challenge and correction of errors.

Westin categories are determined by the responses to the question of whether a respondent agrees with the following three statements. They can be answered on a 5-point Likert scale, with 1 reflecting "complete disagreement" and 5 reflecting "complete agreement". For responses on a 5-point Likert scale, scores of 1 and 2 are classified as disagree, scores of 4 and 5 as agree, and a score of 3 as neutral.

S1. Citizens have lost all control over how personal information about them is circulated and used by companies.

S2. Most businesses handle the personal information they collect about consumers in a proper and confidential way.

S3. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

If respondents agree on S1 and disagree on S2 and S3, then respondents are classified as privacy fundamentalists. Or if respondents agree on S1 and answer neutral on one of S2 and S3, they also classified as privacy fundamentalists. If respondents disagree on S1 and agree on S2 and S3, including a neutral answer either S2 or S3, they are classified as privacy unconcerned. The rest is classified as privacy pragmatists.

4.1.3 Results of Brits and Jonker (2023)

Table 1 provides summary statistics for the benefit and risk scores of using mobile payment apps. The survey gathered responses from 2,465 survey respondents. However, the total number of observations is 2,840, as some people use more than one mobile payment app. The researchers underscore that the median and mean benefit scores of privacy fundamentalists are lower than those of the overall sample. The median and mean benefit scores of privacy unconcerned individuals are greater than those of privacy fundamentalists and privacy pragmatists. The median and mean risk scores of privacy unconcerned individuals are the lowest among the three groups.

Table 1: Summary statistics of benefit and risk scores in Brits and Jonker (2023)

Benefit scores						
Benefit scores by Westin type	Observations	Median	Mean	SD	Min	Max
Payment (all)	2,840	3.33	3.25	1.93	1	5
Privacy fundamentalist	836	3.00	3.03	1.24	1	5
Privacy pragmatist	1,760	3.33	3.32	1.15	1	5
Privacy unconcerned	244	3.67	3.57	1.90	1	5
Risk scores						
Risk scores by Westin type	Observations	Median	Mean	SD	Min	Max
Payment (all)	2.840	3.43	3.41	0.74	1	5
Privacy fundamentalist	836	3.79	3.76	0.72	1.29	5
Privacy pragmatist	1,760	3.29	3.30	0.69	1	5
Privacy unconcerned	244	3.00	2.97	0.67	1	4.86

Note: The table indicates the summary statistics in the survey where respondents were asked to answer benefits and risks regarding the usage of mobile payment apps. Respondents were also classified into three Westin types based on their answers to questions regarding privacy and data usage issues in society. For the definition of each Westin type, refer to Section 4.1.2.

Table 2, based on the actual survey results by Brits and Jonker (2023), illustrates the ratio of perceived benefit versus risk among mobile payment app users and non-users. The table generally reveals a high proportion of both users and non-users exhibiting rational behavior. Specifically, 48.7% of users perceived high benefit and low risk, consistent with rational usage behavior. 46.5% of non-users perceived a low benefit and high risk, which is categorized as rational non-usage behavior. In contrast, the segments reporting paradoxical behavior are relatively small. Users who perceived high risk and low benefit (categorized as paradoxical usage) constituted 10.0% of the group, and non-users perceiving low risk and high benefit (paradoxical non-usage) made up 11.6%.

Table 2: Perceived benefits versus risks of mobile payment apps from Brits and Jonker (2023)

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	48.7%	23.1%	11.6%	12.8%
Low benefit	18.3%	10.0%	29.1%	46.5%

Note: The table reports the proportion of respondents in each group. The threshold value that distinguishes the high benefit group from the low benefit group is the overall median benefit score. Similarly, the overall median risk score is used to distinguish between the high risk group and the low risk group.

4.2 Replication with GenAI

In order to replicate the survey using GenAI, we first assume that the generative agents reside in the Netherlands. Subsequently, we categorize each generative agent into one of six age groups: (1) 15 to 24, (2) 25 to 34, (3) 35 to 44, (4) 45 to 54, (5) 55 to 64, and (6) above 65.[8] Finally, we incorporate the distribution of mobile payment app usage by age category from the actual survey. In each simulation, we randomly designate the generative agent as a user or non-user in the prompt, ensuring that the probability of being a user or non-user aligns for each age group with the actual survey data as the proportion of users varies across different age groups according to their data.[9] The total number of respondents in the actual survey regarding mobile payment apps was 2,405. In contrast, in this study, we simulate the synthetic survey with ChatGPT, aiming to collect 10,000 responses. Table 3 presents the number of responses and the probability of being a mobile payment app user in each age group, reflecting the distribution of each age group in the actual survey.

Table 3: The number of responses and the probability of being a user in each age group

age group	the number of responses	the probability of being a user
from 15 to 24	450	0.6
from 25 to 34	800	0.6
from 35 to 44	1260	0.53
from 45 to 54	1770	0.45
from 55 to 64	1920	0.3
from 65 to 80	3800	0.18

Note: The number of responses in each age group for our experiment reflects the distribution of age groups in the actual survey. The probability of being a user in each age group is determined by the proportion of users in that age group in the actual survey. The overall age distributions in both the synthetic and actual surveys deviate from that of the actual population in the Netherlands.

In order to observe how the sampling of survey respondents and the prompt design affect the overall result of surveys, we conduct three different experimental cases. In Case 1, we

[8]We postulate that the maximum age of a generative agent is 80.
[9]In the actual survey, 35% of total respondents are users and 65% are non-users.

consider the distribution of mobile payment app users along with the age distribution based on the actual survey. In Case 2, we specify the Westin category based on the survey by Brits and Jonker (2023). When we receive responses from each age group, we randomly assign each Westin category in the prompt based on this proportion. In Case 3, we expand the experiment structure of Case 1 to include more features such as income and education level, creating more diverse personas. Unlike in Case 2, we do not assign the Westin category in the prompt randomly but instead rely on survey questions to determine the Westin category.

We show the prompt using the RTF methodology below to implement Case 1. In the first few lines, we detail the role of the generative agent in the prompt. The options for age, and whether the persona is a user or not in the prompt are randomly generated based on their distribution in the actual survey data. {startAge} and {endAge} in instruction B specify the lower and upper boundaries of each age group, and these values are provided by Python programming code. Subsequently, we outline the task of the generative agent, which involves creating a persona and responding to the survey questions. Finally, we instruct the generative agents to respond in a specific format to facilitate efficient parsing of their answers.

A. Choose a random option for gender. [male, female]

B. Choose a random option for the age between {startAge} and {endAge}.

C. Choose a random option for whether you are a user of a mobile payment app or not. In this case, you are going to choose {userChoice}.

Based on the values chosen in steps A-C, create a persona for a survey respondent living in the Netherlands.

As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

{survey questions}[a]

Please give us your answer in the format like this. (the survey question number; scale)

[a]Please refer to the appendix.

We employ ChatGPT-4o, setting the temperature to 1.0 and the top_p to 1.0 across all experimental cases. The temperature parameter modifies the shape of the distribution of potential responses, while the top_p parameter establishes a cutoff point for the cumulative distribution of possible answers. We conduct an analysis and comparison of the results from three distinct experimental cases, utilizing various prompt designs and sampling methodologies.

4.2.1 Case 1: Aligning age group and user distribution with the actual survey

In Case 1, we emulate the user distribution for each age group from the actual survey. We obtain responses from generative agents with a focus on each age group. The decision on whether a generative agent is a user depends on the probability of being a user in each age group. For instance, if we generate responses from the age group ranging from 55 to 64, we designate the generative agent as a user with a 30% probability in instruction C, which aligns with the probability from the actual survey.

Tables 4, 5, and 6 present the overall results of Case 1. The following interesting observations can be drawn from the experiment.

Variation in responses The standard deviations of the benefit and risk scores are smaller than those found in the actual survey. For instance, the standard deviation for benefit scores are substantially lower in the simulation, registering 0.82 (Table 4) compared to 1.93 in the actual survey (Table 1). Similarly, the standard deviation for risk scores is only 0.29 (Table 4), which is significantly smaller than the 0.74 observed in the actual survey (Table 1). Furthermore, the range between the minimum and maximum values of the benefit and risk scores is narrower than that of the actual survey. For example, the maximum benefit score is 4.67 in the simulation (Table 4), compared to 5 in the actual survey (Table 1). The minimum benefit score is 1.33 in the synthetic survey (Table 4), compared to 1 in the actual survey (Table 1).

Westin type The share of privacy fundamentalists is higher than that observed in the actual survey. Specifically, the simulation shows a privacy-fundamentalist proportion of 39.6% (3,961 out of 10,000; Table 4). This is notably higher than the 29.4% observed in the actual survey (836 out of 2,840; Table 1). Conversely, the share of privacy-unconcerned individuals is significantly smaller in the simulation, registering only 0.04% (4 out of 10,000; Table 4), compared with 8.4% (244 out of 2,840; Table 1) in the actual survey. This suggests that generative agents tend to respond to security and privacy-related questions in a more cautious manner on average.[10]

Benefit and risk scores by Westin type Examining the difference across Westin types in Table 4, the benefit scores are likely to decrease and the risk scores tend to increase as agents become more privacy concerned, which is consistent with the actual survey. The median and mean risk scores of the privacy fundamentalist group are the highest among the three groups, and those of the privacy unconcerned group are the lowest. Additionally, the mean of benefit scores for privacy unconcerned agents is the highest while the mean of risk scores is the lowest. It's important to note that we do not specify the negative correlation between views on risk

[10]According to Kumaraguru and Cranor (2005), the privacy fundamentalist group comprises 30 percent, the privacy pragmatist group comprises 60 percent, and the privacy unconcerned people comprises only 10 percent of the whole population.

and benefit in the prompt when we simulate the responses. Therefore, generative agents successfully incorporate into their responses the tendency of privacy concerned people to see more risks while perceiving smaller benefits associated with the apps.

User share by Westin type Table 5 indicates that the share of users in each Westin type is the highest for privacy unconcerned (75%) and lowest for fundamentalist (10.7%), which is align with the actual survey. We also observe that the variation in the share across Westin types is larger in the simulation. For example, the share of users in privacy fundamentalist is the lowest but the difference between fundamentalists and unconcerned is 13 percentage points in the actual survey.

Privacy paradoxical behavior by user type In addition, we observe that the proportion of users and non-users exhibiting privacy paradox behavior is significantly lower than observed in the actual survey, as shown in the top panel of Table 6. Users exhibiting privacy paradoxical behavior have high risk perception and low benefit perception, while non-users exhibiting privacy paradoxical behavior have low risk perception and high benefit perception. This is because the share of users who have benefit scores above the median of all samples is 100%, while that of non-users is substantially lower (9.7% + 8.6% = 18.3%). Put differently, the views on the benefits of financial apps heavily depend on whether the agent is a user or a non-user. This tendency is observed in the actual survey, though it is less pronounced. The share of users who have benefit scores above the median is 71.8%, while that of non-users is 24.4% in the actual survey. In the simulation, the share of non-users whose risk scores are higher than the median (57%) is substantially higher than that of users (11.1%). Again, this risk-averse view of non-users is consistent with the actual survey, where the shares for users and non-users are 33.1% and 59.3%, respectively. We should note that we did not define this difference in characteristics between users and non-users in the prompt. Therefore, this result implies that ChatGPT could provide responses by incorporating the multidimensional aspects of perceptions of actual users and non-users to some extent.

To sum up, without specifying the details of the persona in the prompt, ChatGPT can generate responses that are consistent with the actual survey in terms of the relationship between Westin types and the views on the benefit and risk of payment apps for each type. In addition, it generates responses by successfully incorporating the tendency that users see more benefits and fewer risks in apps than non-users. However, ChatGPT does not provide as much variation as observed in the actual survey. Moreover, generative agents are more likely to be privacy fundamentalist than actual people are.

Table 4: Summary statistics benefit and risk scores for Case 1 with 10,000 responses

Benefit scores						
Westin type	Obs.	Median	Mean	SD	Min	Max
All	10000	2.67	3.07	0.82	1.33	4.67
Privacy fundamentalist	3961	2.33	2.63	0.58	1.33	4.67
Privacy pragmatist	6035	3.67	3.36	0.83	1.33	4.67
Privacy unconcerned	4	3.83	3.67	0.72	2.67	4.33
Risk scores						
Westin type	Obs.	Median	Mean	SD	Min	Max
All	10000	4	3.97	0.29	2.71	4.86
Privacy fundamentalist	3961	4.07	4.08	0.27	3.14	4.86
Privacy pragmatist	6035	3.93	3.9	0.29	2.71	4.86
Privacy unconcerned	4	3.61	3.57	0.27	3.21	3.86

Note: The table indicates the summary statistics in the synthetic survey where 10,000 generative agents in the ChatGPT-4o are asked to answer benefits and risks regarding the usage of mobile payment apps.

Table 5: User ratio by Westin type for Case 1

Westin type	ChatGPT	Bris and Jonker (2023)
All	34.7%	35%
Privacy fundamentalist	10.7%	32%
Privacy Pragmatist	50.4%	35%
Privacy unconcerned	75.0%	45%

Note: This table shows the share of payment app users in each Westin type group, based on the simulation and the actual survey by Bris and Jonker (2023).

4.2.2 Case 2: Specifying Westin categories

In Case 1, we allow generative agents to freely respond to the questions that are used to determine their Westin type without specifying the type in the prompt. As a result, the majority of generative agents are classified as privacy fundamentalist or pragmatists. To reflect the distribution of Westin types in the real population, we specify the Westin type in the prompt by adding Step D. In Step D of the prompt, we assign a specific Westin type along with its description. {westin_type} in the prompt refers to one of the Westin categories: (1) privacy fundamentalist, (2) privacy pragmatist, and (3) privacy unconcerned. {westin_desc} indicates the description of the corresponding Westin category.

Table 6: Perceived benefits and scores for Case 1

All

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	88.8% (48.7%)	11.1% (23.1%)	9.7% (11.6%)	8.6% (12.8%)
Low benefit	0% (18.3%)	0% (10.0%)	33.1% (29.1%)	48.4% (46.5%)

Privacy fundamentalist

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	81.6%	18.3%	6.1%	5.6%
Low benefit	0%	0%	33.7%	54.3%

Privacy pragmatist

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	89.8%	10.1%	14.0%	22.6%
Low benefit	0%	0%	32.2%	41.5%

Privacy Unconcerned

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	100%	0%	0%	0%
Low benefit	0%	0%	100%	0%

Note: The table reports the proportion of generative agents in each group. The threshold value that distinguishes the high benefit group from the low benefit group is the overall median benefit score. Similarly, the overall median risk score is used to distinguish between the high risk group and the low risk group.

A. Choose a random option for gender. [male, female]

B. Choose a random option for the age between {startAge} and {endAge}.

C. The choice for whether you are a user of a mobile payment app or not is "yes" and "no". In this case, you are going to choose {userChoice}.

D. The choices for Westin type are privacy fundamentalist, privacy pragmatist, and privacy unconcerned. In this case, you are going to choose {westin_type}. {westin_desc}

Based on the values chosen in steps A-D, create a persona for a survey respondent living in the Netherlands.

As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

{survey questions}

Please give us your answer in the format like this. (the survey question number; scale)

Table 7 and Table 8 show the results from Case 2. We have three observations worth mentioning. In Table 7, we observe that the medians of benefit scores increase from the privacy fundamentalist group to the privacy unconcerned group, while the medians of risk scores decrease from the privacy fundamentalist group to the privacy unconcerned group, which is also observed in Case 1. However, the difference in the medians or means of risk scores and benefit scores between privacy fundamentalists and privacy unconcerned individuals is more pronounced than in the actual survey and Case 1.

In addition, the standard deviations within each Westin type for benefit and risk scores are smaller than those of the actual survey. However, the standard deviation of risk scores (0.87) for all agents is higher than that of the actual survey (0.74), indicating that the variation of risk scores across different Westin groups is larger compared to the actual survey. This result suggests that specifying Westin type in the prompt has an effect of emphasizing the difference in the preferences of generative agents regarding the risks and benefits of payment apps among Westin types.

In Table 8, the proportions of users and non-users who show privacy paradox behavior are more than doubled compared to the actual survey. This is because the correlation among user/non-user profiles and risk and benefit scores becomes weaker. For example, although the proportion of users who have higher risk scores than the median is 35.2%, more than half of them have lower benefit scores (28.5% of all users). This is contrary to what was observed in the actual survey and Case 1.

Given that the result for Case 1 shows a low proportion with privacy paradox behavior, which is consistent with the actual survey, specifying Westin types in the prompt could be a dominant factor influencing views on risks and benefits by weakening the difference between users and non-users. The comparison between Cases 1 and 2 implies that simultaneously specifying persona in various aspects has a heterogeneous effect on the responses.

4.2.3 Case 3: Adding income and educational level attributes

In Case 1, the simulation results show smaller standard deviations in benefit and risk scores compared to the actual survey. One possible reason for the small variation is that we only specify the attributes of app usage, gender, and age of generative agents in the prompt. In other words, we might have ignored some important attributes that determine views on risk and benefit. Therefore, ChatGPT could be giving responses by assuming average values based on its own model for the attributes that are not specified in the prompt.

The questionnaire of the actual survey included questions regarding the income and ed-

Table 7: Summary statistics benefit and risk scores for Case 2

Group	Observations	Median	Mean	SD	Min	Max
Benefit score						
All	10,000	3.33	3.24	0.78	1	5
Privacy fundamentalist	3,020	2.33	2.37	0.45	1	4
Pragmatist	5,923	3.67	3.50	0.49	2	4.67
Unconcerned	1,057	4.33	4.26	0.48	2.67	5
Risk scores						
Group	**Observations**	**Median**	**Mean**	**SD**	**Min**	**Max**
All	10,000	3.86	3.80	0.87	1.00	5.00
Fundamentalist	3,020	4.71	4.67	0.18	4.07	5.00
Pragmatist	5,923	3.79	3.73	0.23	2.79	4.43
Unconcerned	1,057	1.64	1.70	0.34	1.00	2.86

Note: The table indicates the summary statistics in the synthetic survey where generative agents in the ChatGPT are asked to answer benefits and risks regarding the usage of mobile payment apps. The Westin type of each agent, along with the user or non-user profile, is provided in the prompt. For the definition of each Westin type, refer to Section 4.1.2.

Table 8: Perceived benefits and scores for Case 2

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	62.9% (48.7%)	6.79% (23.1%)	23.4% (11.6%)	6.49% (12.8%)
Low benefit	1.6% (18.3%)	28.5% (10.0%)	25.3% (29.1%)	44.6% (46.5%)

Note: The table reports the proportion of generative agents in each group for the simulations where we specify both the Westin type and the user or non-user profile in the prompt. The threshold value that distinguishes the high benefit group from the low benefit group is the overall median benefit score. Similarly, the overall median risk score is used to distinguish between the high risk group and the low risk group.

ucation level of survey respondents. Researchers observed that demographic factors such as income and education level affected the perceived privacy risk level. For example, Omrani and Soulié (2020) finds that people with higher income tend to perceive more benefits than risks regarding privacy threats. In addition, Bhatia and Breaux (2018) argued that survey respondents with higher education levels perceived less risk.

To investigate this possibility, we elaborate on the prompt design by specifying income and education levels. As shown in the box below, the prompt includes additional features to define a detailed persona. We categorize monthly income into three ranges: (1) EUR 0 to EUR 1,000, (2) EUR 1,001 to EUR 2,000, and (3) EUR 2,001 and above, following the actual survey. For education level, we have two categories: (1) less than a bachelor's degree, and (2) a bachelor's degree or higher.

We assume for simplicity that the distribution of income is independent of the distribution of education. To make Case 3 comparable to Case 1, we specify the persona of the generative agents such that the joint distribution of age and user/non-user attributes of the agents matches the actual distribution.

A. Choose a random option for gender. [male, female]
B. Choose a random option for the age between {startAge} and {endAge}.
C. The choice for the monthly income is ″EUR 0-1000,″ ″EUR 1001-2000,″ and ″EUR 2001 and higher″. In this case, you are going to choose {incomeChoice}.
D. The choice for whether the education level being bachelor or higher is ″yes″ and ″no″. In this case, you are going to choose {eduChoice}.
E. The choice for whether you are a user of a mobile payment app or not is ″yes″ and ″no″. In this case, you are going to choose {userChoice}.
Based on the values chosen in steps A-E, create a persona for a survey respondent living in the Netherlands.
As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

{survey questions}
Please give us your answer in the format like this. (the survey question number; scale)

Table 9: Summary statistics benefit and risk scores for Case 3

Benefit Scores						
Group	**Observations**	**Median**	**Mean**	**SD**	**Min**	**Max**
Total	10,000	2.67	2.95	0.76	1.33	4.67
Fundamentalist	6,103	2.33	2.70	0.63	1.33	4.67
Pragmatist	3,893	3.67	3.36	0.76	1.33	4.67
Unconcerned	4	4.33	4.25	0.17	4.00	4.33
Risk scores						
Group	**Observations**	**Median**	**Mean**	**SD**	**Min**	**Max**
Total	10,000	4.00	4.01	0.28	2.93	4.86
Fundamentalist	6,103	4.07	4.07	0.27	2.93	4.86
Pragmatist	3,893	3.93	3.92	0.28	2.93	4.79
Unconcerned	4	3.82	3.75	0.22	3.43	3.93

Note: The table indicates the summary statistics in the synthetic survey where generative agents in the ChatGPT are asked to answer benefits and risks regarding the usage of mobile payment apps. In the prompt, the income group and education level of agents are specified. The agents were classified into three Westin types based on their answers to questions regarding privacy and data usage issues in society. For the definition of each Westin type, refer to Section 4.1.2.

Table 9 and Table 10 summarize the results of the survey. There are four observations worth mentioning. First, since we do not specify the Westin category in the prompt, the distribution

Table 10: Perceived benefits and scores for Case 3

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	78.0% (48.7%)	21.7% (23.1%)	5.9% (11.6%)	6.6% (12.8%)
Low benefit	0.1% (18.3%)	0.1% (10.0%)	34.8% (29.1%)	52.6% (46.5%)

Note: The table reports the proportion of generative agents in each group for the simulations where the income group and education level of the agents are specified in the prompt. The threshold value that distinguishes the high benefit group from the low benefit group is the overall median benefit score. Similarly, the overall median risk score is used to distinguish between the high risk group and the low risk group.

of Westin categories is similar to that seen in Table 4 for Case 1. Most of the generative agents are classified as "privacy fundamentalists or pragmatist'. We observe that a very small number of generative agents are classified as "privacy unconcerned".

Second, the mean risk score of privacy fundamentalist individuals is the highest among the three groups while their mean benefit score is the lowest among the three groups, which aligns with the actual survey.

Additionally, we observe that the standard deviations of benefit and risk scores for the overall group slightly decreased compared to those in Case 1. Therefore, the standard deviations of benefit and risk scores for the entire sample are also smaller than those from the actual survey.

In Table 10, we can also see that the percentage of privacy paradox behavior from users is nearly zero, and the percentage of privacy paradox behavior from non-users is approximately 6 percent. In addition, the share of people who show rational behavior from users and non-users is higher than those from the actual survey. The overall result of Case 3 appears similar to that of Case 1.

5 Discussion

5.1 Prompt design and the setting of synthetic surveys

We explore potential ways to elaborate and control synthetic survey design regarding the benefit and risk of mobile payment apps by conducting three different experimental cases. Comparing Case 1 and Case 2, we conclude that privacy and benefit perceptions are influenced by specifying Westin categories. In Case 3, we add properties such as income and education level for the generative agents.

To investigate how the design of synthetic surveys affects the overall benefit and risk scores, we conduct t-tests on the average benefit and risk scores from the three different experimental cases. Each cell in Table 11 contains the t-statistic value for the test of the difference in aver-

Table 11: The Welch's t-test results among three experimental cases

benefit	Case 1	Case 2	Case 3
Case 1		15.0***	10.0***
Case 2			26.6***
Case 3			

risk	Case 1	Case 2	Case 3
Case 1		18.5***	22.9***
Case 2			22.9***
Case 3			

Note: The table reports t-statistics for the test on the difference in average scores between two cases. (***, **, and * stand for statistical significance at the 0.001, 0.01, and 0.05 level respectively after Bonferroni correction.

age scores between cases. Interestingly, the difference in benefit or risk scores is statistically significant for all pairs after the Bonferroni correction. This suggests that the prompt design substantially affects responses.

These results provide a simple lesson for changing prompt design. We should note that when we modified the prompt design, our goal was not to change the average benefit or risk scores. However, the prompt generated a significantly different result for the those scores although the difference is small. This implies that we cannot necessarily obtain the expected results by modifying prompts and should be aware that it may lead to unintended consequences.

There are some limitations on our prompt design approach. First, while we aim to match the marginal distributions of individual characteristics (e.g., age, income, gender) in our generative agents to those of the target population, the strategy of randomly varying items in the prompt may inadvertently produce combinations that violate real-world correlations.

Second, generative agents are fundamentally modeled as a collection of specified attributes (age, gender, income, Westin type, etc.). We must acknowledge that human respondents are more than a mere summation of these characteristics; they possess nuanced personalities, situational context, and free will, which are challenging or impossible to capture within a concise set of prompt attributes.

5.2 Temperature and variation in responses

In Case 1, we observe small variations in the credit and benefit scores compared to the actual survey. In Case 3, we provide specific attributes for educational level and income, aiming to generate some variation. Even with these variations in attributes, the standard deviation of the scores remains low, although it increased slightly in Case 3.

Another way to generate variation in responses is to adjust the value of the temperature parameter, which was set to 1.0 in all three cases. To examine whether increasing the temperature helps to generate enough variation to mimic the actual distribution, we simulate the survey by adjusting the temperature from 0.6 to 1.4 for 1,000 times. The standard deviations of benefit

Table 12: Standard deviations with different temperature

temperature	Benefit	Risk
1.4	0.81	0.33
1.2	0.82	0.32
1	0.82	0.29
0.8	0.82	0.29
0.6	0.81	0.27
Brits and Jonker (2023)	1.93	0.74

Note: The table indicates the standard deviations of risk and benefit scores in the simulations with different temperature values.

and risk scores in each simulation are shown in Table 12. The table shows that increasing the temperature parameter from 0.6 to 1.4 does not increase the variation in the benefit score. On the other hand, we found some differences in the variance of the risk score among the results with different temperature values. The standard deviation increases from 0.27 to 0.33 by 20%. However, even with a temperature parameter of 1.4, the standard deviations of benefit scores from the simulation is much smaller than that of the actual survey. We generated the responses with a temperature of 1.5 to obtain more variation as an extreme case. However, the proportion of responses with a temperature of 1.5 that are not valid becomes higher.[11]

These results indicate a caveat of using ChatGPT for marketing surveys, especially from the viewpoint of policy makers. Policy makers, such as central banks, do not consider only the opinions of median or average respondents. The same principle applies to decision makers in the private sector. Therefore, generating enough variation observed in the actual survey and incorporating minority opinions is of paramount importance. Our experiments indicate that policy makers need to exercise great caution when utilizing ChatGPT for market surveys, as it could ignore the tails of the distribution.

5.3 Generative agents as privacy fundamentalists and sources

of the bias As shown in Case 1, the share of generative agents who responded as privacy fundamentalists is higher than that of the actual survey. This means that the proportion of generative agents who agree with the statement that "citizens have lost all control over how personal information about them is circulated and used by companies" is higher. This is one example of bias that AI can generate.

There are many possible reasons why ChatGPT generates this bias, as pointed out by Ferrara (2023), among many others. In addition, defining and measuring biases in LLMs provokes many

[11]We see invalid answers for survey questions, such as blank responses and text entries.

related issues. As Ferrara (2023) argues, LLMs such as ChatGPT merely reflect society, which contains various biases, stereotypes, and assumptions. Furthermore, it is not a straightforward task to define what "bias" means. For example, cultural norms can vary across time and regions. If we consider the bias of results based on "fairness," the bias could be subjective. Therefore, stakeholders of AI models should define what "fairness" means in their context.[12] Despite these challenges, all stakeholders should continue to cooperate toward reducing bias. In particular, if policy makers use GenAI to develop surveys on consumers' perceptions of risk and privacy, they should recognize the risk of bias.

To examine the sensitivity of bias to the location of generative agents, we simulated another survey with a slightly altered prompt, where the only change was the resident country of the generative agents, from the Netherlands to the United States. If ChatGPT's responses are influenced by training data that carry different biases across countries and if ChatGPT takes into account the residential country specified in the prompt, we would expect the bias observed in the responses to vary when the resident country of the generative agents changes. By specifying the resident country as the United States, 73% of generative agents are categorized as privacy fundamentalists, 25.8% as pragmatists, and 1.2% as unconcerned. This finding was particularly surprising, in contrast to surveys showing that respondents in the Netherlands are less willing to share data than respondents in the United States (Chen et al. (2023), Vitak et al. (2023)).This result suggests that the bias in our experiment does not appear to depend on the residence country of the generative agents.

One possible source of this bias is the adjustment made through reinforcement learning from human feedback. To minimize the risk of AI responding to questions in a way that disregards privacy or produces toxic outputs, ChatGPT is fine-tuned using reinforcement learning from human feedback.[13] This fine-tuning could result in a high proportion of privacy-conscious generative agents. Strictly speaking, it would not be possible for ChatGPT users to formally test or determine the exact reasons behind biased responses. However, the additional experiment indicates that recognizing and eliminating biases is a challenging task.

5.4 Synthetic survey with ChatGPT-4.0

Other language models may exhibit different tendencies in their responses. To explore this possibility, we conducted a simulation using ChatGPT-4.0 with the same prompt as in Case 1. Table 13 presents the summary statistics of risk and benefit scores by Westin type. Interestingly, the proportion of privacy-unconcerned agents is again very low, which contrasts with the

[12]See Blodgett et al. (2020) for example.
[13]For example, see Ouyang et al. (2022) for details.

nearly 10% share observed in the real survey. Furthermore, the pragmatists' share is lower than the results with ChatGPT-4o shown in Table 4. These findings suggest that generative agents created by ChatGPT-4.0 tend to be more concerned about privacy than those generated by ChatGPT-4o.

In addition, the same pattern observed with ChatGPT-4o emerges: privacy fundamentalists exhibit higher risk scores while having lower benefit scores compared to privacy pragmatists. The standard deviation of benefit and risk scores are higher than those from ChatGPT-4o. However, their levels are lower than those from the actual survey.

Finally, Table 14 reports the results for the privacy calculus. As in the simulation with ChatGPT-4o, whether a generative agent is a user or non-user significantly affects benefit scores. Approximately 99% of users have benefit scores higher than the median across all samples, while 80.6% of non-users have lower benefit scores. In summary, ChatGPT-4.0 generates responses with patterns similar to those from ChatGPT-4o.

Table 13: Summary statistics benefit and risk scores for Case 1 using ChatGPT-4.0

Benefit scores						
Benefit scores by Westin type	Observations	Median	Mean	SD	Min	Max
Payment (all)	1,000	2.33	2.92	1.15	1	5
Privacy fundamentalist	495	2.33	2.43	0.89	1	4.66
Privacy pragmatist	486	3.66	3.37	1.17	1	5
Privacy unconcerned	19	4.33	4.19	0.67	2	4.66
Risk scores						
Risk scores by Westin type	Observations	Median	Mean	SD	Min	Max
Privacy (all)	1,000	3.71	3.69	0.49	1	4.85
Privacy fundamentalist	495	3.85	3.79	0.475	1.286	4.85
Privacy pragmatist	486	3.64	3.59	0.49	1	4.78
Privacy unconcerned	19	3.28	3.26	0.39	2.64	3.92

Note: The table indicates the summary statistics in the synthetic survey where generative agents in the ChatGPT-4.0 are asked to answer benefits and risks regarding the usage of mobile payment apps.

Table 14: Perceived benefits and scores for Case 1 using ChatGPT-4.0

	Users		Non-users	
	Low risk	High risk	Low risk	High risk
High benefit	61.9% (48.7%)	37.7% (23.1%)	8.5% (11.6%)	10.7% (12.8%)
Low benefit	0.0% (18.3%)	0.3% (10.0%)	27.3% (29.1%)	53.3% (46.5%)

Note: The table reports the proportion of generative agents in each group for the simulations. The threshold value that distinguishes the high benefit group from the low benefit group is the overall median benefit score. Similarly, the overall median risk score is used to distinguish between the high risk group and the low risk group.

6 Conclusion

This paper has explored the potential of GenAI for conducting a survey about payment methods, especially in terms of user perception of benefits and privacy risks. We select the actual survey conducted by Brits and Jonker (2023), and conduct different experimental cases. We find that the responses from GenAI are consistent with the privacy calculus tendency observed in the actual survey for payment app users. Additionally, generative agents categorized as privacy unconcerned have more favorable views on the benefits of financial apps without specifying those characteristics in prompts, which aligns with the actual survey. Moreover, ChatGPT can replicate the tendency that the views of generative agents on the benefits of financial apps are more widely dispersed than those on risks. By conducting different sets of experiments, we find that the prompt design and the setting of synthetic surveys affect the overall results of surveys.

However, our experiments suggest some caveats when ChatGPT is used for a synthetic survey. First, it should be noted that the results from GenAI do not provide the same level of variation observed in the actual survey. Specifying detailed personas in the prompt and increasing the value of the temperature parameter to increase randomness resulted in minor changes in the standard deviations of benefit and risk scores. The low variation could lead to misleading conclusions by ignoring minority views, which could be a fatal error for policy makers.

Another caveat is that most generative agents are classified as "privacy fundamentalists," which is not consistent with the actual survey. In other words, ChatGPT generates a bias in views on privacy and risks related to payment apps. To examine the source of this bias, we simulate another experiment by changing the residence country of generative agents from the Netherlands to the United States and obtained an even higher share of privacy fundamentalists. This suggests that the source of the bias is difficult to examine and fix. In other words, users of GenAI should recognize the possibility of biases and the difficulty of eliminating such biases.

Finally, by specifying the Westin type in the prompt, the difference between users and non-users becomes less clear in terms of risk and benefit scores, although the variation across different Westin types becomes more pronounced. This result suggests that changing prompts leads to unexpected consequences and could force ChatGPT to put less weight on some features and more weight on others, even if both are specified in the prompt.

Our findings suggest that GenAI has the potential to be used as a copilot tool to conduct surveys on marketing related to payment apps. GenAI could increase productivity in designing surveys. However, our results indicate that GenAI cannot completely replace surveys conducted with actual human beings. Due diligence is necessary to identify and cultivate GenAI use cases while carefully managing associated risks.

References

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Bhatia, J. and Breaux, T. D. (2018). Empirical measurement of perceived privacy risk. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(6):1–47.

Binz, M. and Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., and Larson, J. M. (2024). Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, 32(4):401–416.

Blodgett, S. L., Barocas, S., Daumé III, H., and Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *arXiv preprint arXiv:2005.14050*.

Brand, J., Israeli, A., and Ngwe, D. (2023). Using GPT for market research. *Available at SSRN 4395751*.

Brits, H. and Jonker, N. (2023). The use of financial apps: Privacy paradox or privacy calculus? *De Nederlandsche Bank Working Paper No. 794*.

Chen, L., Huang, Y., Ouyang, S., and Xiong, W. (2022). Data privacy and digital demand. Technical report, Tech. rep., National Bureau of Economic Research.

Chen, S., Doerr, S., Frost, J., Gambacorta, L., and Shin, H. S. (2023). The fintech gender gap. *Journal of Financial Intermediation*, 54:101026.

Choi, J. H., Hickman, K. E., Monahan, A. B., and Schwarcz, D. (2021). ChatGPT goes to law school. *J. Legal Educ.*, 71:387.

Cornelli, G., Frost, J., and Mishra, S. (2023). Artificial intelligence, services globalisation and income inequality. Technical report, Bank for International Settlements.

ECB (2023). A stocktake on the digital euro.

Ferrara, E. (2023). Should ChatGPT be biased? challenges and risks of bias in large language models. *arXiv preprint arXiv:2304.03738*.

Ghaffarzadegan, N., Majumdar, A., Williams, R., and Hosseinichimeh, N. (2023). Generative agent-based modeling: Unveiling social system dynamics through coupling mechanistic models with generative artificial intelligence. *arXiv preprint arXiv:2309.11456*.

Gimpel, H., Kleindienst, D., and Waldmann, D. (2018). The disclosure of private data: measuring the privacy paradox in digital services. *Electronic Markets*, 28(4):475–490.

Goldfarb, A. and Que, V. F. (2023). The economics of digital privacy. *Annual Review of Economics*, 15(1):267–286.

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., and Steinhardt, J. (2020). Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Higham, C. F., Higham, D. J., and Grindrod, P. (2023). Diffusion models for generative artificial intelligence: An introduction for applied mathematicians. *arXiv preprint arXiv:2312.14977*.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn form homo silicus? *National Bureau of Economic Research Working Paper No. 31122*.

Kazinnik, S. (2023). Bank run, interrupted: Modeling deposit withdrawals with generative AI. *Available at SSRN: https://ssrn.com/abstract=4656722*.

Kumaraguru, P. and Cranor, L. F. (2005). Privacy indexes: a survey of westin's studies.

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., et al. (2023). Performance of ChatGPT on USMLE: potential for ai-assissted medical education using large language models. *PLoS digital health*, 2(2):e0000198.

Li, J. (2023). Predicting the demand for central bank digital currency: A structural analysis with survey data. *Journal of Monetary Economics*, 134:73–85.

Li, P., Castelo, N., Katona, Z., and Sarvary, M. (2023). Language models for automated market research: A new way to generate perceptual maps. *Available at SSRN 4241291*.

Ludwig, J., Mullainathan, S., and Rambachan, A. (2025). Large language models: An applied econometric framework. National Bureau of Economic Research Working Paper.

Ma, D., Zhang, T., and Saunders, M. (2023). Is ChatGPT humanly irrational?

Motoki, F., Monteiro, J., Malagueño, R., and Rodrigues, V. (2023). Charting new territory: Using chatgpt to enhance survey instruments in the organizational setting. *Available at SSRN: https://ssrn.com/abstract=4595896*.

Omrani, N. and Soulié, N. (2020). Privacy experience, privacy perception, political ideology and online privacy concern: the case of data collection in europe. *Revue d'économie industrielle*, (172):217–255.

OpenAI (2024). Gpt-4o system card. `https://openai.com/index/gpt-4o-system-card/`. System Card published by OpenAI on August 8, 2024. Accessed on 4 September 2025.

OpenAI (2025). Model Release Notes: Updates to GPT-4o in ChatGPT. `https://help.openai.com/en/articles/9624314-model-release-notes`. Accessed on 4 September 2025.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22.

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., and Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*.

Perez-Cruz, F. and Shin, H. S. (2024). Testing the cognitive limits of large language models. *BIS Bulletin No 83*.

Vitak, J., Liao, Y. Y., Mols, A., Trottier, D., Zimmer, M., Kumar, P., and Pridmore, J. (2023). When do data collection and use become a matter of concern?: A cross-cultural comparison of us and dutch privacy attitudes. *International journal of communication (Online)*, 17(2023):471–498.

Williams, M., Nurse, J. R., and Creese, S. (2016). The perfect storm: The privacy paradox and the internet-of-things. In *2016 11th International Conference on Availability, Reliability and Security (ARES)*, pages 644–652. IEEE.

Appendix

Survey questions

This section demonstrates the survey questions and explains how the risk and benefit scores were calculated in Brits and Jonker (2023). Table A1 contains survey questions regarding data protection and data privacy. Similar to Brits and Jonker (2023), we ask generative agents on the severity and the likelihood of the inappropriate use of personal data. Each generative agent can answer the privacy sensitivity using a 5 point Likert scale, with 1 reflecting "not at all threatening" and 5 reflecting "extremely threatening". This is called the severity risk. Generative agents are also asked about the likelihood that an incident related to inappropriate data handling would happen in 12 months. This is the likelihood risk. A 5 point Likert scale, with 1 reflecting "very unlikely" and 5 reflecting "extremely threatening" was used to calculate the likelihood risk. The subjective risk is the arithmetic average of the answers to 7 questions.

Subjective_risks = $\sum_{j=1}^{7}(0.5 \times Serverity\_risks\_j + 0.5 \times Likelihood\_risks\_j)$

In addition to risk scores, we use a set of questions to calculate benefit scores.

1. I think that using a mobile payment app is useful in my daily live;
2. it will be easy to use the mobile payment app.

Table A1: The survey questions regarding risk perception (Brits and Jonker (2023)

Description inappropriate data handling	In short
1. The company uses my personal information to ask me a higher price for a product than others because it sees which products I find attractive and how much I am willing to pay for them	Price discrimination
2. The company uses my personal information to have me make impulse purchases through enticing advertisements	Impulse purchases
3. The company uses my personal information without my knowledge for anything other than what I have consented to	Inappropriate data use
4. The company sells my personal data to another company, without my knowledge	Data sale
5. The company passes my personal data to government agencies, without my knowledge	Government
6. Employees of the company peek into my personal data without my permission	Inappropriate data access staff
7. People outside the company can access my personal data if the company is hacked or due to data breaches	Data hack / breach

3. it can be joyful to use the mobile payment app.

The synthetic agents show their level of agreement on these three statements using a 5 Likert scale with 1 reflecting "complete disagreement" and 5 reflecting "complete agreement". The unweighted arithmetic mean of the answers to three questions is used for the benefit score.

Responses from ChatGPT

Figure A1 illustrates typical responses obtained in the simulation with ChatGPT.

```
1 1;Male, 2;21, 3;No, 4;4, 5;3, 6;2, 7;3, 8;2, 9;4, 10;5, 11;4, 12;3, 13;5, 14;3, 15;2, 16;4, 17;5,
18;4, 19;4, 20;5, 21;2, 22;3, 23;3
2 1;Male, 2;20, 3;No, 4;3, 5;3, 6;4, 7;3, 8;2, 9;4, 10;4, 11;4, 12;4, 13;3, 14;3, 15;2, 16;4, 17;4,
18;3, 19;4, 20;3, 21;2, 22;3, 23;2
3 1;Male, 2;20, 3;No, 4;4, 5;3, 6;2, 7;3, 8;4, 9;5, 10;5, 11;4, 12;3, 13;5, 14;3, 15;4, 16;5, 17;5,
18;4, 19;3, 20;5, 21;3, 22;3, 23;4
4 1;Female, 2;23, 3;No, 4;4, 5;2, 6;3, 7;2, 8;3, 9;4, 10;4, 11;3, 12;2, 13;4, 14;2, 15;3, 16;4,
17;3, 18;3, 19;2, 20;3, 21;3, 22;4, 23;5
5 1;Male, 2;19, 3;No, 4;5, 5;2, 6;2, 7;4, 8;3, 9;5, 10;5, 11;4, 12;5, 13;5, 14;3, 15;3, 16;4, 17;5,
18;4, 19;5, 20;5, 21;3, 22;3, 23;3
6 1;Male, 2;20, 3;Yes, 4;3, 5;3, 6;2, 7;4, 8;4, 9;3, 10;5, 11;3, 12;2, 13;3, 14;2, 15;3, 16;3, 17;2
18;2, 19;1, 20;4, 21;5, 22;4, 23;5
7 1;Male, 2;20, 3;Yes, 4;3, 5;4, 6;2, 7;3, 8;3, 9;4, 10;2, 11;2, 12;3, 13;4, 14;2, 15;2, 16;3, 17;1
18;2, 19;3, 20;4, 21;5, 22;4, 23;4
8 1;Female, 2;20, 3;No, 4;4, 5;2, 6;3, 7;4, 8;3, 9;5, 10;5, 11;4, 12;4, 13;3, 14;4, 15;3, 16;4,
17;5, 18;5, 19;4, 20;3, 21;2, 22;3, 23;2
9 1;Female, 2;22, 3;Yes, 4;3, 5;4, 6;2, 7;3, 8;4, 9;3, 10;2, 11;1, 12;2, 13;3, 14;2, 15;3, 16;3,
17;1, 18;1, 19;2, 20;3, 21;5, 22;5, 23;4
```

Figure A1: Sample responses from ChatGPT-4.o

Note: The figure shows the sample responses from ChatGPT-4.o. As stated in the prompt, ChatGPT-4.o returns answers to survey questions in the format (the question number;answer). Semicolons are used as delimiters.

Programming code

The box below shows Python programming code used for the research. The code used the API service provided by OpenAI.

```python
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
import time
import pandas as pd
from openai import OpenAI
from dotenv import load_dotenv
import os
import random
load_dotenv()
# Define constants
sleep_time = 0.3
length_per_iter = 1

# Include API key provided by OpenAI service
env_api_key = os.environ['OPENAI_API_KEY']
client = OpenAI(api_key=env_api_key, timeout=20.0)

# All functions for prompting, extracting responses, etc.
def query(prompt):
    rvec = []
    response = client.chat.completions.create(
        model="gpt-4",
        messages=[{"role": "system", "content": prompt}],
        temperature=1.0,
        max_tokens=1000,
        top_p=1.0,
        n=1,
        frequency_penalty=0.0,
        presence_penalty=0.0
    )
    rvec.append(response)
    return rvec

def westin_category():
```

```python
    rand_num = random.randrange(1, 11)
    westin_type = ""
    westin_desc = ""
    if rand_num <= 3:
        westin_type = "Privacy Fundamentalist"
        westin_desc = "You are a privacy fundamentalist, who consider privacy very important and do not believe \
        your personal data are in safe hands when shared"
    elif rand_num >= 4 and rand_num <= 9:
        westin_type = "Privacy Pragmatist"
        westin_desc = "You are a privacy pragmatist, who do have some concerns but in the same time do have a certain \
        amount of trust that your personal data are handled properly."
    elif rand_num == 10:
        westin_type = "Privacy Unconcerned"
        westin_desc = "You are a privacy unconcerned, who are hardly concerned about your privacy and have little \
        problems sharing their data."
    return westin_type, westin_desc

def demo_config(from_age=25, to_age=34):
    num = 0
    age = 0
    gender = "female"
    age = random.randrange(from_age, to_age+1)
    num = random.randrange(1, 101)

    if num >= 1 and num <= 49: # 49% of population older than 12
        gender = "male"
    else: # 51% of population older than 12
        gender = "female"
    return str(age), gender

def get_user_option(user_prob):
    prob = random.random()
    if prob <= user_prob:
        return "yes"
    return "no"
```

```python

def get_monthly_income(incomeProb1, incomeProb2, incomeProb3):
    income_choice = random.random()
    if income_choice <= incomeProb1:
        return "EUR 0 and EUR 1000"
    elif income_choice > incomeProb1 and income_choice <= (incomeProb1 + incomeProb2):
        return "EUR 1001 and EUR 2000"
    elif income_choice > (1 - (incomeProb1 + incomeProb2)):
        return "EUR 2001 and above"
    return None

def get_education(eduProb1, eduProb2):
    edu_choice = random.random()
    if edu_choice <= eduProb1:
        return "less than bachelor"
    elif edu_choice > (1 - eduProb1):
        return "bachelor or higher"
    return None

def get_region():
    region_choice = random.randrange(1, 5)
    if region_choice == 1:
        return "West"
    elif region_choice == 2:
        return "East"
    elif region_choice == 3:
        return "North"
    elif region_choice == 4:
        return "South"
    return None

def make_prompt_case1(startAge="15", endAge="24", userChoice="yes", country="Netherlands"):
    prompt = f"""
    A. Choose a random option for gender. [male, female]
    B. Choose a random option for the age between {startAge} and {endAge}.
```

```
C. The choice for whether you are a user of a mobile payment app or not is "yes" or "no".
In this case, you are going to choose {userChoice}.

Based on the values chosen in steps A-C, create a persona for a survey respondent living in {country}. Please do not print the choice and the description of each persona.

As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

1. What is your gender? (male, female)
2. What is your age?
3. Do you use a mobile payment app? (yes, no)

Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
On a 5 point Likert scale ranging from 1(complete disagreement) to 5(complete agreement),
to what extent do you agree with the following statements?

4. Citizens have lost all control over how personal information about them is circulated and used by companies.
5. Most businesses handle the personal information they collect about consumers in a proper and confidential way.
6. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
On a 5 point Likert scale ranging from 1 \'not at all threatening\' to 5 \'extremely threatening\', how threatening would you consider possible privacy breaches assuming you are using a mobile payment app?

7. The company uses my personal information to ask me a higher price for a product than
```

```
   others because it sees which products I find attractive and how much I am willing to pay for them.
8. The company uses my personal information to have me make impulse purchases through enticing advertisements.
9. The company uses my personal information without my knowledge for anything other than what I have consented to.
10. The company sells my personal data to another company, without my knowledge.
11. The company passes my personal data to government agencies, without my knowledge.
12. Employees of the company peek into my personal data without my permission.
13. People outside the company can access my personal data if the company is hacked or due to data breaches.

Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
Below you will see a number of possible privacy-related incidents that could happen use of mobile payment apps.
On a 5 point Likert scale ranging from 1 \'very unlikely\' to 5 \'very likely\',
please indicate how likely you think it is that such an incident will happen to you in the next 12 months assuming that you are using a mobile payment app.

14. The company uses my personal information to ask me a higher price for a product than
    others because it sees which products I find attractive and how much I am willing to pay for them.
15. The company uses my personal information to have me make impulse purchases through enticing advertisements.
16. The company uses my personal information without my knowledge for anything other than what I have consented to.
17. The company sells my personal data to another company, without my knowledge.
18. The company passes my personal data to government agencies, without my knowledge.
```

```
    19. Employees of the company peek into my personal data without my
    ↪ permission.
    20. People outside the company can access my personal data if the company is
    ↪ hacked or due to data breaches.

    Please provide a range of perspectives that can be useful for understanding
    ↪ various user experiences and attitudes towards mobile payment apps.
    On a 5 point Likert scale ranging from 1 (complete disagreement) to 5
    ↪ (complete agreement), to what extent do you agree with these if you think
    ↪ of an 'imaginary' app usage.

    21. I think that using a mobile payment app is useful in my daily life.
    22. It will be easy to use the mobile payment app.
    23. It can be joyful to use the mobile payment app.

    Please give me your answer in the format like this (the question
    ↪ number;answer). And please put comma(,) between answers. Please do not
    ↪ tell me any reason.
    Respondent: """
    return prompt

def make_prompt_case2(startAge="15", endAge="24", userChoice="yes",
↪ country="Netherlands", westin_type="Privacy Fundamentalist", westin_desc=""):

    prompt = f"""
    A. Choose a random option for gender. [male, female]
    B. Choose a random option for the age between {startAge} and {endAge}.
    C. The choice for whether you are a user of a mobile payment app or not is
    ↪ "yes" or "no".
       In this case, you are going to choose {userChoice}.
    D. The choices for Westin type are privacy fundamentalist, privacy
    ↪ pragmatist, and privacy unconcerned.
       In this ase, you are going to choose {westin_type}. {westin_desc}

    Based on the values chosen in steps A-D, create a persona for a survey
    ↪ respondent living in {country}. Please do not print the choice and the
    ↪ description of each persona.

```

```
As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

1. What is your gender? (male, female)
2. What is your age?
3. Do you use a mobile payment app? (yes, no)
4. What is your choice for Westin type categories? (fundamentalist, pragmatist, unconcerned)

Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
On a 5 point Likert scale ranging from 1 \'not at all threatening\' to 5 \'extremely threatening\', how threatening would you consider possible privacy breaches assuming you are using a mobile payment app?

5. The company uses my personal information to ask me a higher price for a product than
   others because it sees which products I find attractive and how much I am willing to pay for them.
6. The company uses my personal information to have me make impulse purchases through enticing advertisements.
7. The company uses my personal information without my knowledge for anything other than what I have consented to.
8. The company sells my personal data to another company, without my knowledge.
9. The company passes my personal data to government agencies, without my knowledge.
10. Employees of the company peek into my personal data without my permission.
11. People outside the company can access my personal data if the company is hacked or due to data breaches.

Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
Below you will see a number of possible privacy-related incidents that could happen use of mobile payment apps.
```

```
    On a 5 point Likert scale ranging from 1 \'very unlikely\' to 5 \'very likely\',
    please indicate how likely you think it is that such an incident will happen to you in the next 12 months assuming that you are using a mobile payment app.

    12. The company uses my personal information to ask me a higher price for a product than
        others because it sees which products I find attractive and how much I am willing to pay for them.
    13. The company uses my personal information to have me make impulse purchases through enticing advertisements.
    14. The company uses my personal information without my knowledge for anything other than what I have consented to.
    15. The company sells my personal data to another company, without my knowledge.
    16. The company passes my personal data to government agencies, without my knowledge.
    17. Employees of the company peek into my personal data without my permission.
    18. People outside the company can access my personal data if the company is hacked or due to data breaches.

    Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
    On a 5 point Likert scale ranging from 1 (complete disagreement) to 5 (complete agreement), to what extent do you agree with these if you think of an 'imaginary' app usage.

    19. I think that using a mobile payment app is useful in my daily life.
    20. It will be easy to use the mobile payment app.
    21. It can be joyful to use the mobile payment app.



    Please give me your answer in the format like this (the question number;answer). And please put comma(,) between answers. Please do not tell me any reason.
    Respondent: """
```

```python
    return prompt


def make_prompt_case3(startAge="15", endAge="24", incomeChoice="EUR 0-1000", eduChoice="no", userChoice="no", country="Netherlands"):
    prompt = f"""
    A. Choose a random option for gender. [male, female]
    B. Choose a random option for the age between {startAge} and {endAge}.
    C. The choice for the monthly income is "EUR 0-1000", "EUR 1001-2000", and "EUR 2001 and higher".
    In this case, you are going to choose {incomeChoice}.
    D. The choice for whether the education level being bachelor or higher is "yes" and "no".
    In this case, you are going to choose {eduChoice}.
    E. The choice for whether you are a user of a mobile payment app or not is "yes" and "no".
    In this case, you are going to choose {userChoice}.

    Based on the values chosen in steps A-E, create a persona for a survey respondent living in {country}. Please do not print the choice and the description of each persona.

    As a survey respondent based on the persona, please fill out this form, providing various perspectives on the benefits and risks of using mobile payment apps.

    1. What is your gender? (male, female)
    2. What is your age?
    3. What is your monthly income? (EUR 0-1000, EUR 1001-2000, EUR 2001 and higher)
    4. What is your education level equals to bachlor or higher? (no, yes)
    5. Do you use a mobile payment app? (no, yes)

    Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
    On a 5 point Likert scale ranging from 1(complete disagreement) to 5(complete agreement),
    to what extent do you agree with the following statements?
```

```

    6. Citizens have lost all control over how personal information about them is circulated and used by companies.
    7. Most businesses handle the personal information they collect about consumers in a proper and confidential way.
    8. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

    Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
    On a 5 point Likert scale ranging from 1 \'not at all threatening\' to 5 \'extremely threatening\', how threatening would you consider possible privacy breaches assuming you are using a mobile payment app?

    9. The company uses my personal information to ask me a higher price for a product than
       others because it sees which products I find attractive and how much I am willing to pay for them.
    10. The company uses my personal information to have me make impulse purchases through enticing advertisements.
    11. The company uses my personal information without my knowledge for anything other than what I have consented to.
    12. The company sells my personal data to another company, without my knowledge.
    13. The company passes my personal data to government agencies, without my knowledge.
    14. Employees of the company peek into my personal data without my permission.
    15. People outside the company can access my personal data if the company is hacked or due to data breaches.

    Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
    Below you will see a number of possible privacy-related incidents that could happen use of mobile payment apps.
    On a 5 point Likert scale ranging from 1 \'very unlikely\' to 5 \'very likely\',
```

```
    please indicate how likely you think it is that such an incident will happen to you in the next 12 months assuming that you are using a mobile payment app.

    16. The company uses my personal information to ask me a higher price for a product than
        others because it sees which products I find attractive and how much I am willing to pay for them.
    17. The company uses my personal information to have me make impulse purchases through enticing advertisements.
    18. The company uses my personal information without my knowledge for anything other than what I have consented to.
    19. The company sells my personal data to another company, without my knowledge.
    20. The company passes my personal data to government agencies, without my knowledge.
    21. Employees of the company peek into my personal data without my permission.
    22. People outside the company can access my personal data if the company is hacked or due to data breaches.

    Please provide a range of perspectives that can be useful for understanding various user experiences and attitudes towards mobile payment apps.
    On a 5 point Likert scale ranging from 1 (complete disagreement) to 5 (complete agreement), to what extent do you agree with these if you think of an 'imaginary' app usage.

    23. I think that using a mobile payment app is useful in my daily life.
    24. It will be easy to use the mobile payment app.
    25. It can be joyful to use the mobile payment app.

    Please give me your answer in the format like this (the question number;answer). And please put comma(,) between answers. Please do not tell me any reason.
    Respondent: """
    return prompt

def write_to_excel(total_answer, sheet_idx):
```

```python
    result_df = pd.DataFrame(total_answer, columns=['agent-id', 'choice'])

    with pd.ExcelWriter(file_path, mode='w') as writer:
        result_df.to_excel(writer, sheet_name="trial_{0}".format(sheet_idx),
         ↪  engine="xlsxwriter")

age_group_list = [(15, 24), (25, 34), (35, 44), (45, 54), (55, 64), (65, 80)]
user_prob_list = [0.6, 0.6, 0.53, 0.45, 0.3, 0.18]
user_responses_list = [50, 100, 150, 200, 200, 400]

# age_group from 15 to 24 -> index 0
# age_group from 25 to 34 -> index 1
# age_group from 35 to 44 -> index 2
# age_group from 45 to 54 -> index 3
# age_group from 55 to 64 -> index 4
# age_group from 65 to 80 -> index 5

experiment_case = 1
age_group_idx = 1

N = user_responses_list[age_group_idx]
from_age = age_group_list[age_group_idx][0]
to_age = age_group_list[age_group_idx][1]

income_prob1 = 0.74
income_prob2 = 0.22
income_prob3 = 0.04

edu_prob1 = 0.75
edu_prob2 = 0.25

user_prob = user_prob_list[age_group_idx]
file_path = "./v4.0_t1.0_%d_%d_1.xlsx" % (from_age, to_age)

result_df = None
total_answer = []
index = 0

```

```
for i in range(0, N):
    outputline = []
    age, gender = demo_config(from_age=from_age, to_age=to_age)
    app = "None"
    is_user = 'Y'

    user_choice = get_user_option(user_prob)
    income_choice = get_monthly_income(income_prob1, income_prob2, income_prob3)
    edu_choice = get_education(edu_prob1, edu_prob2)
    westin_type, westin_desc = westin_category()
    if westin_type == "Privacy Fundamentalist":
        user_prob = 0.32
    elif westin_type == "Privacy Pragmatist":
        user_prob = 0.35
    elif westin_type == "Privacy Unconcerned":
        user_prob = 0.45

    user_choice = get_user_option(user_prob)

    if experiment_case == 1:
        result_prompt = make_prompt_case1(from_age, to_age, user_choice,
         ↪  "Netherlands")
    elif experiment_case == 2:
        result_prompt = make_prompt_case2(from_age, to_age, user_choice,
         ↪  "Netherlands", westin_type, westin_desc)
    elif experiment_case == 3:
        result_prompt = make_prompt_case3(from_age, to_age, income_choice,
         ↪  edu_choice, user_choice, "Netherlands")

    responses = query(result_prompt)
    choices = responses[0]
    temp = choices.choices
    choice = temp[0].message.content
    choice_cleaned = choice.replace('\n','')
    choice_cleaned = choice_cleaned.replace('.', '')
    temp[0].message.content + "\n"
    outputline.append(i)
    outputline.append(choice_cleaned)
```

```
    print(str(i) + " " + choice_cleaned)
    total_answer.append(outputline)
    time.sleep(sleep_time)

write_to_excel(total_answer, index)
```

Previous volumes in this series

1332 February 2026	Dollar funding and housing markets: the role of non-US global banks	Torsten Ehlers, Mathias Hoffmann and Alexander Raabe
1331 February 2026	The perils of narrowing fiscal spaces	Hanno Kase, Leonardo Melosi, Sebastian Rast and Matthias Rottner
1330 February 2026	Passive investors and loan spreads	Konrad Adler, Sebastian Doerr and Sonya Zhu
1329 February 2026	Adoption and welfare effects of payment innovations: the case of digital wallets in Peru	Arturo Andia, Jose Aurazo and Marcelo Paliza
1328 February 2026	The perils of narrowing fiscal spaces	Hanno Kase, Leonardo Melosi, Sebastian Rast and Matthias Rottner
1327 January 2026	A tractable menu cost model with an aggregate markup drift	Ko Munakata
1326 January 2026	Monetary policy and private equity acquisitions: tracing the links	Fernando Avalos, Boris Hofmann and Jose Maria Serena
1325 January 2026	AI adoption, productivity and employment: evidence from European firms	Iñaki Aldasoro, Leonardo Gambacorta, Rozalia Pal, Debora Revoltella, Christoph Weiss and Marcin Wolfski
1324 January 2026	The macro-financial impact of economic policy uncertainty in Latin America	Ana Aguilar, Rafael Guerra, Carola Müller and Alexandre Tombini
1323 January 2026	Banks and capital requirements: evidence from countercyclical buffers	Iñaki Aldasoro, Andreas Barth, Laura Comino Suarez and Riccardo Reale
1322 January 2026	Embracing gen AI: a comparison of Italian and US households	David Loschiavo, Olivier Armantier, Antonio Dalla Zuanna, Leonardo Gambacorta, Mirko Moscatelli and Ilaria Supino
1321 December 2025	Artificial intelligence and growth in advanced and emerging economies: short-run impact	Leonardo Gambacorta, Enisse Kharroubi, Aaron Mehrotra and Tommaso Oliviero