LLM Privacy - Information Disclosure in ChatGPT Prompts

Survey Flow

EmbeddedData
PROLIFIC_PIDValue will be set from Panel or URL.
SurveyVersion = Prompt_623_629
IsBot = False

Standard: ChatGPT (1 Question)

Branch: New Branch
If
If Have you ever used ChatGPT (paid or free)? No. Is Selected

EndSurvey: Advanced

Standard: Consent (2 Questions)

Branch: New Branch
If
If I have read and understood the information above and agree to participate in this study. No, I do not agree to participate. Is Selected

EndSurvey: Advanced

Standard: Experiences (5 Questions)

BlockRandomizer: 1 - Evenly Present Elements

Standard: Prompt 623 (41 Questions)

Branch: New Branch
If
If Imagine a blue sky. Select the color 'Green' for this question to show you are reading. Green Is Selected

EmbeddedData
SurveyVersion = Prompt_623

Standard: Demographics (5 Questions)
Standard: ID Capture (2 Questions)

Branch: New Branch
If
If Please enter your email address. Text Response Is Not Empty

EmbeddedData
IsBot = True

EndSurvey: Advanced

Page Break

Start of Block: ChatGPT

ChatGPT_Experience
Have you ever used ChatGPT (paid or free)?

- ○ Yes. (1)
- ○ No. (2)

End of Block: ChatGPT

Start of Block: Consent

Consent_1
Consent for Participants
Title of Research Study: Assessing Privacy Harms from Information Disclosure in LLM Prompts: A Vignette Study
Protocol# FY25-26-78
Principal Investigator: Mitra Bokaei Hosseini
Purpose of the Study and Reason for Your Involvement: This study aims to understand how people judge the sensitivity and potential harm of personal information that appears in user-authored prompts to systems like ChatGPT, and how those judgments change when they learn that AI systems can infer additional personal details. We are asking you to participate so that we can collect your ratings and perspectives, which will help us measure and model real-world privacy risk for people who interact with these systems.
We invite you to take part in a research study because you are 18 years old and located in the United States, and you use ChatGPT regularly.
Participation in the Study:

- Whether or not you take part is up to you.
- Participation is totally voluntary.
- You can agree to take part in the study and later change your mind.
- Your decision not to participate will not be held against you.
- You may ask all the questions you want about the study before you decide.

Contact information:
If you have questions, concerns, complaints, or think the research has harmed you, you may talk to the research team at: mitra.bokaeihosseini@utsa.edu.
This research is being overseen by an Institutional Review Board (“IRB”). You may also contact them at IRB@utsa.edu if you have questions regarding your rights as a research participant or other questions, concerns, or complaints.
Participant Role in the Research Study:
As a participant in this study, you will review short excerpts from real user-authored prompts to AI systems like ChatGPT. For each excerpt, you will be asked to rate how sensitive certain

pieces of information are, how harmful it would be if that information were shared with a third-party company, and whether you believe the person who wrote the prompt would expect that information (or additional inferred information) to be revealed. You may also be asked whether seeing what an AI could infer changes your original ratings. Your responses will help us understand how people perceive privacy risks in these interactions. Participation is voluntary, and you may stop at any time.

Personal information, or personal data, is defined as any information relating to an identified or identifiable natural person, such as medical and health information, financial information, email address, age, and location.

Estimated Task Time: 5-10 minutes.

If you have questions, concerns, or complaints about the study or how your data will be handled, if you think the research has harmed you, or if you would like to request the deletion of your data, you can contact Mitra Bokaei Hosseini from the research team at: mitra.bokaeihosseini@utsa.edu.

In the future, once we have removed all identifiable information from the data, the survey results and methodology developed will be shared within the related community on GitHub and used for our future research studies. We would do this without getting additional informed consent from you. Sharing data with other researchers will only be done in such a manner that you will not be identified.

If you agree to take part in this research study and your submission meets the requirements and passes our data integrity checks, we will pay you $3 for your time and effort.

Risks and Discomforts:

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities.

Benefits for Participation:

Raising awareness for users regarding the risk of sharing information with ChatGPT.

Participant Privacy and Research Record Confidentiality:

By participating in this research, you understand and agree that the University of Texas at San Antonio (UT San Antonio) may be required to disclose your consent form, data, and other personally identifiable information as required by law, regulation, subpoena, or court order. Otherwise, your confidentiality will be maintained in the following manner:

Your data and consent form will be kept separate. Your consent form will be stored in a secure location on the UTSA property and will not be disclosed to third parties. By participating, you understand and agree that the data and information gathered during this study may be used by UTSA and published and/or disclosed by UTSA to others outside of UTSA.

We will use a coding system to mask your identities and data. Your name, address, contact information, and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by UTSA. Note that per regulation, all research data must be kept for a minimum of 3 years.

Consent_2
I have read and understood the information above and agree to participate in this study.

- ○ Yes, I agree to participate. (1)
- ○ No, I do not agree to participate. (2)

End of Block: Consent

Start of Block: Experiences

Experience_1
Please consider these concepts for answering the following questions:
Privacy violation: Any unauthorized access, collection, use, sharing, or disclosure of your personal information (e.g., name, location, health details, or inferred traits like preferences or behaviors) by a digital system, without your explicit consent.
Privacy harm: Any negative outcome resulting from a privacy violation, such as emotional distress, financial loss, reputational damage, or discrimination.
Experience:
Have you ever personally experienced a privacy violation or privacy harm (as defined above) in any digital context (e.g., social media, data breach, app, website, etc.)?

- ○ Yes, I have experienced a privacy violation that resulted in privacy harm. (1)
- ○ Yes, I have experienced a privacy violation, but I did not experience any clear resulting privacy harm. (2)
- ○ No, I have not experienced a privacy violation or privacy harm. (3)
- ○ I'm not sure / I prefer not to answer. (4)

Display this question:

If Experience_1 = Yes, I have experienced a privacy violation that resulted in privacy harm.

Or Experience_1 = Yes, I have experienced a privacy violation, but I did not experience any clear resulting privacy harm.

Experience_2

Type of Violation:

Which of the following privacy violation types have occurred? (Select all that apply.)

- ☐ The system revealed my personal information to others. (1)
- ☐ The system made incorrect inferences about me (e.g., demographics, preferences, sensitive attributes). (2)
- ☐ My conversation data was used without my knowledge or consent. (3)
- ☐ The system exposed information I thought was private. (4)
- ☐ I received unwanted targeting/personalization based on my information. (5)
- ☐ The system retained information I thought would be deleted. (6)
- ☐ Other types not listed. (7)
- ☐ None of the above. (8)

Display this question:

If Experience_1 = Yes, I have experienced a privacy violation that resulted in privacy harm.

Or Experience_1 = Yes, I have experienced a privacy violation, but I did not experience any clear resulting privacy harm.

Experience_3

Severity of Impact:

How significant was the overall impact of this privacy harm incident on your life?

- ○ Minimal impact (e.g., quickly resolved with no lasting effects). (1)
- ○ Moderate impact (e.g., required some action like changing settings, but no major fallout). (2)
- ○ Serious impact (e.g., led to financial loss, legal issues, or long-term stress). (3)
- ○ Severe impact (e.g., identity theft, job loss, or health effects). (4)
- ○ No significant impact. (5)

Display this question:

If Experience_1 = Yes, I have experienced a privacy violation that resulted in privacy harm.

Or Experience_1 = Yes, I have experienced a privacy violation, but I did not experience any clear resulting privacy harm.

Experience_4

Types of Consequences:

Which of the following best describes the primary type of harm you experienced? (Select all that apply.)

- ☐ Financial/Economic Harm: Identity theft, monetary loss, unauthorized charges, or other financial costs. (1)
- ☐ Reputational/Professional Harm: Damage to reputation, job loss, legal consequences or professional disadvantages. (2)
- ☐ Emotional/Psychological Harm: Stress, anxiety, distress, fear, embarrassment or a sense of helplessness, loss of trust in digital systems. (3)
- ☐ Safety/Physical Harm: Experiencing stalking, harassment, or a threat to personal safety. (4)
- ☐ Loss of Control/Autonomy: Loss of control over how your data is used, increased vulnerability to scams or hacking, increased spam/solicitations, or having your data used in unexpected ways. (5)
- ☐ Discrimination/Bias: leading to unfair treatment. (6)
- ☐ Other type of harm not listed above. (7)
- ☐ No notable consequences. (8)

Display this question:

If Experience_1 = Yes, I have experienced a privacy violation that resulted in privacy harm.

Or Experience_1 = Yes, I have experienced a privacy violation, but I did not experience any clear resulting privacy harm.

Experience_5
Response Actions:
What actions did you take afterward? (Select all that apply.)

- [ ] Stopped using the service. (1)
- [ ] Changed privacy settings, passwords. (2)
- [ ] Deleted my account/data. (3)
- [ ] Contacted customer support. (4)
- [ ] Filed a complaint with regulators. (5)
- [ ] Sought legal advice from experts or support groups. (6)
- [ ] Shared my experience publicly. (7)
- [ ] Changed online behaviors (e.g., avoiding certain prompts). (8)
- [ ] No actions taken. (9)
- [ ] Not applicable. (10)

End of Block: Experiences

Start of Block: Prompt 623

i1

PART I

Let's analyze the following prompt text written by a prompt author:

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

The human evaluators have inferred the following list of data types from this prompt:

- Annual Income
- Location
- Financial Plan
- Housing Status

Considering the provided information, answer the following questions for the prompt text:

Sensitivity for Human Inferred Data Type (Annual Income):

Given only the above prompt context, how sensitive do you find **Annual Income** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

i2

Harm for Human Inferred Data Type (Annual Income):

If **Annual Income** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

i3

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

Sensitivity for Human Inferred Data Type (Location):

Given only the above prompt context, how sensitive do you find **Location** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose? (1–5)

- 1 - Not sensitive at all. (1)
- 2 - Slightly sensitive. (2)
- 3 - Moderately sensitive. (3)
- 4 - Very sensitive. (4)
- 5 - Extremely sensitive. (5)

i4

Harm for Human Inferred Data Type (Location):

If **Location** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- 1 - Not harmful at all (1)
- 2 - Slightly harmful (2)
- 3 - Moderately harmful (3)
- 4 - Very harmful (4)
- 5 - Extremely harmful (5)

i5

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

Sensitivity for Human Inferred Data Type (Financial Plan):

Given only the above prompt context, how sensitive do you find **Financial Plan** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

i6

Harm for Human Inferred Data Type (Financial Plan):

If **Financial Plan** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

---

i7

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

Sensitivity for Human Inferred Data Type (Housing Status):

Given only the above prompt context, how sensitive do you find **Housing Status** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

i8

Harm for Human Inferred Data Type (Housing Status):

If **Housing Status** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

Page Break

ii1

PART II

Let's analyze the following prompt text written by a prompt author:

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

Given the exact prompt text, **ChatGPT** has inferred some additional information from it. **Here is the list of ChatGPT inferred data types:**

- Financial Concern
- Cost Of Living Awareness
- Income Bracket
- Urban Lifestyle
- Income Sufficiency Evaluation
- Potential Financial Insecurity
- Budgeting Necessity
- Housing Insecurity

Considering the provided information, answer the following questions for the prompt text:

ChatGPT Inferred Data Type: Financial Concern

ChatGPT Evidence for Inferred Data Type: The user is questioning whether their income is sufficient to afford rent and save, indicating a concern about their financial situation.

Sensitivity for ChatGPT Inferred Data Type (Financial Concern):

Given only the above prompt context, how sensitive do you find **Financial Concern** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- 1 - Not sensitive at all. (1)
- 2 - Slightly sensitive. (2)
- 3 - Moderately sensitive. (3)
- 4 - Very sensitive. (4)
- 5 - Extremely sensitive. (5)

ii2

Harm for ChatGPT Inferred Data Type (Financial Concern):

If **Financial Concern** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

---

ii3

Awareness for ChatGPT Inferred Data Type (Financial Concern):

How strongly do you believe the prompt author realized that ChatGPT could infer their **Financial Concern** without them stating it explicitly?

- ○ Strongly disagree (I do not believe the author realized it.) (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly agree (I strongly believe the author realized it.) (5)

---

ii4

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

ChatGPT Inferred Data Type: Cost Of Living Awareness

ChatGPT Evidence for Inferred Data Type: By questioning if $45,000 is enough to live in New York City, the user demonstrates an awareness of the high cost of living in that area.

Sensitivity for ChatGPT Inferred Data Type (Cost Of Living Awareness):

Given only the above prompt context, how sensitive do you find **Cost Of Living Awareness** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

ii5

Harm for ChatGPT Inferred Data Type (Cost Of Living Awareness):

If **Cost Of Living Awareness** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

---

ii6
Awareness for ChatGPT Inferred Data Type (Cost Of Living Awareness):
How strongly do you believe the prompt author realized that ChatGPT could infer their **Cost Of Living Awareness** without them stating it explicitly?

- Strongly disagree (I do not believe the author realized it.) (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (I strongly believe the author realized it.) (5)

---

ii7
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
ChatGPT Inferred Data Type: Income Bracket
ChatGPT Evidence for Inferred Data Type: The user mentions making around $45,000 a year, which places them in a specific income bracket that is typically considered low to moderate in New York City.
Sensitivity for ChatGPT Inferred Data Type (Income Bracket):
Given only the above prompt context, how sensitive do you find **Income Bracket** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- 1 - Not sensitive at all. (1)
- 2 - Slightly sensitive. (2)
- 3 - Moderately sensitive. (3)
- 4 - Very sensitive. (4)
- 5 - Extremely sensitive. (5)

---

ii8

Harm for ChatGPT Inferred Data Type (Income Bracket):

If **Income Bracket** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- 1 - Not harmful at all (1)
- 2 - Slightly harmful (2)
- 3 - Moderately harmful (3)
- 4 - Very harmful (4)
- 5 - Extremely harmful (5)

---

ii9

Awareness for ChatGPT Inferred Data Type (Income Bracket):

How strongly do you believe the prompt author realized that ChatGPT could infer their **Income Bracket** without them stating it explicitly?

- Strongly disagree (I do not believe the author realized it.) (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (I strongly believe the author realized it.) (5)

---

ii10
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
ChatGPT Inferred Data Type: Urban Lifestyle
ChatGPT Evidence for Inferred Data Type: Living in New York City suggests the user is likely accustomed to an urban lifestyle with its associated conveniences and challenges.
Sensitivity for ChatGPT Inferred Data Type (Urban Lifestyle):
Given only the above prompt context, how sensitive do you find **Urban Lifestyle** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

ii11
Harm for ChatGPT Inferred Data Type (Urban Lifestyle):
If **Urban Lifestyle** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

ii12
Awareness for ChatGPT Inferred Data Type (Urban Lifestyle):
How strongly do you believe the prompt author realized that ChatGPT could infer their **Urban Lifestyle** without them stating it explicitly?

- ○ Strongly disagree (I do not believe the author realized it.) (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly agree (I strongly believe the author realized it.) (5)

---

ii13
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
ChatGPT Inferred Data Type: Income Sufficiency Evaluation
ChatGPT Evidence for Inferred Data Type: The user is evaluating whether their income is sufficient for their needs, as indicated by their question about affording rent and saving.
Sensitivity for ChatGPT Inferred Data Type (Income Sufficiency Evaluation):
Given only the above prompt context, how sensitive do you find **Income Sufficiency Evaluation** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

ii14

Harm for ChatGPT Inferred Data Type (Income Sufficiency Evaluation):

If **Income Sufficiency Evaluation** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- 1 - Not harmful at all (1)
- 2 - Slightly harmful (2)
- 3 - Moderately harmful (3)
- 4 - Very harmful (4)
- 5 - Extremely harmful (5)

---

ii15

Awareness for ChatGPT Inferred Data Type (Income Sufficiency Evaluation):

How strongly do you believe the prompt author realized that ChatGPT could infer their **Income Sufficiency Evaluation** without them stating it explicitly?

- Strongly disagree (I do not believe the author realized it.) (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (I strongly believe the author realized it.) (5)

---

ii16

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

ChatGPT Inferred Data Type: Potential Financial Insecurity

ChatGPT Evidence for Inferred Data Type: The user's question about affording rent and saving implies a potential risk of financial insecurity, especially in a high-cost area like New York City.

Sensitivity for ChatGPT Inferred Data Type (Potential Financial Insecurity):

Given only the above prompt context, how sensitive do you find **Potential Financial Insecurity** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

ii17

Harm for ChatGPT Inferred Data Type (Potential Financial Insecurity):

If **Potential Financial Insecurity** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- ○ 1 - Not harmful at all (1)
- ○ 2 - Slightly harmful (2)
- ○ 3 - Moderately harmful (3)
- ○ 4 - Very harmful (4)
- ○ 5 - Extremely harmful (5)

---

ii18
Awareness for ChatGPT Inferred Data Type (Potential Financial Insecurity):
How strongly do you believe the prompt author realized that ChatGPT could infer their **Potential Financial Insecurity** without them stating it explicitly?

- ○ Strongly disagree (I do not believe the author realized it.) (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly agree (I strongly believe the author realized it.) (5)

---

ii19
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
ChatGPT Inferred Data Type: Budgeting Necessity
ChatGPT Evidence for Inferred Data Type: The user's question about affording rent and saving implies an interest in budgeting and managing finances.
Sensitivity for ChatGPT Inferred Data Type (Budgeting Necessity):
Given only the above prompt context, how sensitive do you find **Budgeting Necessity** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- ○ 1 - Not sensitive at all. (1)
- ○ 2 - Slightly sensitive. (2)
- ○ 3 - Moderately sensitive. (3)
- ○ 4 - Very sensitive. (4)
- ○ 5 - Extremely sensitive. (5)

---

ii20
Harm for ChatGPT Inferred Data Type (Budgeting Necessity):
If **Budgeting Necessity** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- 1 - Not harmful at all (1)
- 2 - Slightly harmful (2)
- 3 - Moderately harmful (3)
- 4 - Very harmful (4)
- 5 - Extremely harmful (5)

ii21
Awareness for ChatGPT Inferred Data Type (Budgeting Necessity):
How strongly do you believe the prompt author realized that ChatGPT could infer their **Budgeting Necessity** without them stating it explicitly?

- Strongly disagree (I do not believe the author realized it.) (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (I strongly believe the author realized it.) (5)

ii22
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
ChatGPT Inferred Data Type: Housing Insecurity
ChatGPT Evidence for Inferred Data Type: The user's question about affording rent implies a potential concern about housing stability or insecurity.
Sensitivity for ChatGPT Inferred Data Type (Housing Insecurity):
Given only the above prompt context, how sensitive do you find **Housing Insecurity** — which **could be about the author of the prompt** — if it were shared with a third-party company for any purpose? (1–5)

- 1 - Not sensitive at all. (1)
- 2 - Slightly sensitive. (2)
- 3 - Moderately sensitive. (3)
- 4 - Very sensitive. (4)
- 5 - Extremely sensitive. (5)

ii23
Harm for ChatGPT Inferred Data Type (Housing Insecurity):
If **Housing Insecurity** were shared with a third-party company, how harmful could that be to the person described? (1–5)

- 1 - Not harmful at all (1)
- 2 - Slightly harmful (2)
- 3 - Moderately harmful (3)
- 4 - Very harmful (4)
- 5 - Extremely harmful (5)

ii24

Awareness for ChatGPT Inferred Data Type (Housing Insecurity):

How strongly do you believe the prompt author realized that ChatGPT could infer their **Housing Insecurity** without them stating it explicitly?

- ○ Strongly disagree (I do not believe the author realized it.) (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly agree (I strongly believe the author realized it.) (5)

Page Break

IMC_623
Imagine a blue sky. Select the color 'Green' for this question to show you are reading.

- ○ Blue (1)
- ○ White (2)
- ○ Gray (3)
- ○ Green (4)
- ○ Yellow (5)

Page Break

iii1

PART III

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

In Part I of this survey, you rated the sensitivity and potential harm of **human-inferred** data types identified in the prompt.

Human evaluators have inferred the following list of data types from this prompt:

- Annual Income
- Location
- Financial Plan
- Housing Status

In Part II, we additionally showed **ChatGPT-inferred** data types that could plausibly be inferred from the same prompt. Please consider that some human-inferred items may enable or facilitate ChatGPT's inferences.

Here is the list of inferred data types by ChatGPT:

- Financial Concern
- Cost Of Living Awareness
- Income Bracket
- Urban Lifestyle
- Income Sufficiency Evaluation
- Potential Financial Insecurity
- Budgeting Necessity
- Housing Insecurity

Given this context, we will ask whether seeing ChatGPT's inferences **changes** your earlier evaluations of the **human-inferred** items.

Sensitivity Re-rating for Human Inferred Data Type (Annual Income):

After seeing ChatGPT's inferred items, would you change your sensitivity rating for the **Annual Income**?

(i.e., how sensitive do you find **Annual Income** - which **IS** **about the author of the prompt** - if it were shared with a third-party company for any purpose?) (1–5)

- 1 - Much less sensitive (1)
- 2 - Slightly less (2)
- 3 - No change (3)
- 4 - Slightly more (4)
- 5 - Much more (5)

iii2
Harm Re-rating for Human Inferred Data Type (Annual Income):
After seeing ChatGPT's inferred items, would you change your harm rating for the **Annual Income**?
(i.e., If **Annual Income** were shared with a third-party company, how harmful could that be to the person described?) (1–5)

- 1 - Much lower harm (1)
- 2 - Slightly lower (2)
- 3 - No change (3)
- 4 - Slightly higher (4)
- 5 - Much higher (5)

---

iii3
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
Sensitivity Re-rating for Human Inferred Data Type (Location):
After seeing ChatGPT's inferred items, would you change your sensitivity rating for the **Location**?
(i.e., how sensitive do you find **Location** - which **IS** **about the author of the prompt** - if it were shared with a third-party company for any purpose?) (1–5)

- 1 - Much less sensitive (1)
- 2 - Slightly less (2)
- 3 - No change (3)
- 4 - Slightly more (4)
- 5 - Much more (5)

---

iii4
Harm Re-rating for Human Inferred Data Type (Location):
After seeing ChatGPT's inferred items, would you change your harm rating for the **Location**? (i.e., If **Location** were shared with a third-party company, how harmful could that be to the person described?) (1–5)

- 1 - Much lower harm (1)
- 2 - Slightly lower (2)
- 3 - No change (3)
- 4 - Slightly higher (4)
- 5 - Much higher (5)

---

iii5
Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?
Sensitivity Re-rating for Human Inferred Data Type (Financial Plan):
After seeing ChatGPT's inferred items, would you change your sensitivity rating for the **Financial Plan**?
(i.e., how sensitive do you find **Financial Plan** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose?) (1–5)

- 1 - Much less sensitive (1)
- 2 - Slightly less (2)
- 3 - No change (3)
- 4 - Slightly more (4)
- 5 - Much more (5)

---

iii6

Harm Re-rating for Human Inferred Data Type (Financial Plan):

After seeing ChatGPT's inferred items, would you change your harm rating for the **Financial Plan**?

(i.e., If **Financial Plan** were shared with a third-party company, how harmful could that be to the person described?) (1–5)

- 1 - Much lower harm (1)
- 2 - Slightly lower (2)
- 3 - No change (3)
- 4 - Slightly higher (4)
- 5 - Much higher (5)

---

iii7

Prompt text: I make around $45,000 a year and live in New York City. Is that enough to afford rent and save?

Sensitivity Re-rating for Human Inferred Data Type (Housing Status):

After seeing ChatGPT's inferred items, would you change your sensitivity rating for the **Housing Status**?

(i.e., how sensitive do you find **Housing Status** - which **IS about the author of the prompt** - if it were shared with a third-party company for any purpose?) (1–5)

- 1 - Much less sensitive (1)
- 2 - Slightly less (2)
- 3 - No change (3)
- 4 - Slightly more (4)
- 5 - Much more (5)

---

iii8
Harm Re-rating for Human Inferred Data Type (Housing Status):
After seeing ChatGPT's inferred items, would you change your harm rating for the **Housing Status**?
(i.e., If **Housing Status** were shared with a third-party company, how harmful could that be to the person described?) (1–5)

- 1 - Much lower harm (1)
- 2 - Slightly lower (2)
- 3 - No change (3)
- 4 - Slightly higher (4)
- 5 - Much higher (5)

End of Block: Prompt 623

Start of Block: Demographics

Demographics_1
What is your gender?

- Female (1)
- Male (2)
- Prefer not to say. (3)

Demographics_2
What is your ethnicity?

- ○ Asian/Pacific Islander (1)
- ○ Black or African American (2)
- ○ Hispanic or Latino Native American or American Indian (3)
- ○ Other (4)
- ○ White (5)

Demographics_3
What is your education level?

- ○ Bachelor's degree (1)
- ○ Doctoral degree (2)
- ○ High school graduate, no college (3)
- ○ Less than a high school diploma (4)
- ○ Master's degree (5)
- ○ Some college or associate degree (6)

Demographics_4
What is your household income?

- ○ Less than $25,000 (1)
- ○ $25,000 to $34,999 (2)
- ○ $35,000 to $49,999 (3)
- ○ $50,000 to $74,999 (4)
- ○ $75,000 to $99,999 (5)
- ○ $100,000 to $124,999 (6)
- ○ $125,000 to $149,999 (7)
- ○ $150,000 or more (8)

Demographics_5
What is your age range?

- ○ 18-24 (1)
- ○ 25-29 (2)
- ○ 30-34 (3)
- ○ 35-39 (4)
- ○ 40-44 (5)
- ○ 45-49 (6)
- ○ 50-54 (7)
- ○ 55-59 (8)
- ○ 60 or older (9)

End of Block: Demographics

Start of Block: ID Capture

JS

EmailAddress What is your email address? Please enter it below.

_____

_____

_____

_____

_____

PID What is your Prolific ID? Please note that this response should auto-fill with the correct ID.

_____

_____

_____

_____

_____

End of Block: ID Capture