

# Synthetic User Research in Industry

JIE LI

MARCH 4, 2024





Jie Li

Jie sounds like letter "J"

2009-2021  

MSc in Industrial Design Engineering, TU Delft

PhD in Human-Computer Interaction, TU Delft

Postdoctoral Researcher at CWI

Social VR Experience Design and Measurement

2021-2022 

Principal Researcher, EPAM

2022-Now 

Head of Research & Insights (NL), EPAM



jieli-research.com





2017 - Now

Owner & Cake Artist

Cake Researcher

Delft, Netherlands



cake-researcher.com



@cake\_researcher





2023 - now

## ACM Interactions

- Writing a column entitled “Bits to Bites”
- Reflecting on HCI methodologies in academic and commercial contexts



## Jie Li

The 1994 movie *Eat Drink Man Woman*, directed by Ang Lee, tells the story of widowed master chef Mr. Chu, who expresses his love for his family by preparing feasts every Sunday. During their last family meal, Mr. Chu shares a memorable speech: “Life is not like the mise en place before cooking. Life is spontaneous. You will never be fully prepared for it.”

This speech deeply resonated with me. As someone inclined toward



perfectionism, I often find myself meticulously planning, fearing unexpected situations, and spending excessive time

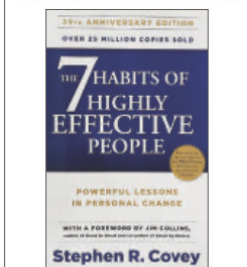
in preparation without taking action. What’s worse is that even when I am finally prepared, I may discover that my passion has diminished. Over the years, I’ve learned the importance of being proactive and realizing that seemingly insignificant tasks can lead to transformative changes.

Recently, while commuting, I’ve been absorbed in the audiobook of Stephen Covey’s *The 7 Habits of Highly Effective People*. Initially, I was wary of books centered on productivity and self-management, but I decided to give it a chance as a non-native speaker of English looking for a convenient listen. To my surprise, I became captivated by the content and connected with the habits and real-life examples presented. Two habits that particularly caught my attention were “Be Proactive” and “Begin with the End in Mind.” Although these habits sound simple, implementing them in my daily life has required significant effort to overcome perfectionism and procrastination.

As both an HCI/UX researcher and a pastry chef, my constant challenge is to

ensure research rigor and deliver satisfactory cake outcomes despite facing limitations in time and equipment. These constraints have taught me to approach problems with a different perspective and find creative solutions.

Drawing parallels with HCI research, the Wizard



of Oz research technique stands out. It involves manually operating the system behind the scenes while users interact with it, unaware of the orchestrated experience. This method allows for gathering valuable insights without the need for a fully functional prototype and demonstrates that a lack of resources does not have to hinder the achievement of research goals, such as evaluating a technology that requires expensive hardware or does not yet exist.

Similarly, I recall my student days when I lacked specialized baking tools and had to seek substitutes and adapt. Although I didn’t need a Wizard per se, I had to get

creative in making cakes when I couldn’t attain the right equipment. For years, a metal mooncake box served as my baking tray, and I adjusted the temperature of my old, overly temperamental oven by gauging how much I should open the oven door. If a particular baking mold was unavailable, I explored alternatives like 3D printing to create it myself. I minimized the belief that I couldn’t accomplish something due to lacking specific tools. In short, two crucial habits are “Be Proactive” and “Begin with the End in Mind.”

My name is Jie Li. I am an HCI/UX researcher and the owner of a boutique café in the Netherlands. Beginning in the January–February 2024 issue, I will be hosting a column in *ACM Interactions* called Bits to Bites, where I will discuss various HCI research methods used in academia, industry, and diverse sectors such as fashion, food, automotive, and immersive technology.

• Jie Li is head of research and insights at EPAM Netherlands. She has a Ph.D. in human-computer interaction from Delft University of Technology. She is also a cake artist and owner of the boutique café Cake Researcher. ✉ [jjasminejue@gmail.com](mailto:jjasminejue@gmail.com)

DOI: 10.1145/3626966 COPYRIGHT HELD BY AUTHOR

INTERACTIONS.ACM.ORG

©INTERACTIONS MAG

INTERACTIONS JANUARY–FEBRUARY 2024 VOLUME XXXI.1

Association for Computing Machinery



Running A/B Testing in Corporate Environments  
Mobile Analytics from Situated Knowledge  
Evaluating Interpretive Research in HCI  
Conversations in Sketch: Keeping Things Real with Peter Kariuki

### COLUMN BITS TO BITES

Jie Li, EPAM

## Experimentation Everywhere and Every Day

Running A/B Testing in Corporate Environments

Imagine opening a webpage to book hotel rooms and transportation for your next holiday. As you click through pages and make selections, you might not realize that you’re using just one of the tens of different variations of the webpage. You’re unknowingly participating in randomized controlled experiments that contribute to companies improving your holiday planning experiences.

The fitness app you’re using could be testing two notification strategies: You and the friend you’re traveling with might receive different types of notifications—one emphasizing competition and another focusing on personal progress. With thousands of users contributing to data collection while using the app, the app company gains insights into which strategy triggers higher user engagement and adherence.

It’s not just digital products; even your local supermarkets and department stores run tests. For instance, they might rearrange popular products at the entrance or at the back or change the direction of escalators to evaluate which variation optimizes customer shopping flows or generates more sales.

### THE ROOT OF EXPERIMENTATION

The foundation of such experimentation is rooted in randomized controlled trials (RCTs) from a clinical background, which is a standard procedure for assessing whether a cause-and-effect relationship exists between medical treatments and outcomes [1]. In the basic format of RCTs, subjects—often patients—are

randomly assigned to one of two groups. One group is the experimental group, where subjects receive an intervention or a new type of treatment. The other group is the control group, which receives an alternative or conventional treatment. Researchers follow up with both groups of subjects and observe whether there are differences in the outcomes.

Inheriting from RCTs, researchers in human-computer interaction also conduct controlled experiments to gain insights into human behavior, cognitive processes, physiological reactions, and interactions with new technology. Unlike RCTs, where subjects are typically patients, participants in HCI experiments are usually drawn from the general population or belong to a specific target user group for the technology under development. In contrast to RCTs, which focus on the effects of specific treatments on patients, HCI experiments often involve a broader range of independent variables (e.g., visual stimuli, devices, and interface design) and dependent variables (e.g., psychological and behavioral measurements). These measurements encompass subjective data, such as self-report questionnaires, as well as more-objective metrics like task-completion time, accuracy rates, physiological sensor data, and qualitative data, including

researchers’ observations and post-experiment interviews [2].

As a specific type of controlled experiment, A/B testing is widely used in the corporate context for digital product development and user experience optimization. It typically involves well-defined A and B variations and a specific set of metrics (e.g., click-through rates, conversion rates) to measure user behavior. In today’s business landscape, nearly all major companies—especially those focused on digital products and software—utilize A/B testing at scale [3]. They conduct hundreds of A/B tests on millions of users daily, covering a wide range of products, including Web platforms, mobile applications, and content arrangements for marketing campaigns. The data collected from these ongoing tests empowers companies to make data-driven decisions, rather than relying solely on the opinions of the highest-paid person.

**THINGS TO BE AWARE OF WITH (AUTOMATED) A/B TESTING**

Although the idea behind A/B testing is straightforward, it can become tedious when a company frequently tests a wide range of prototypes and products. Some big tech companies have their in-house platforms automate the A/B testing process, including traffic allocation (i.e., the decision on how users are divided between different variations), statistical analysis, and result interpretation. This automation helps reduce manual effort and speeds up decision making (e.g., [4]). Companies that can’t afford in-house solutions can leverage other commercial platforms such as VWO (<https://vwo.com>).

**You’re unknowingly participating in randomized controlled experiments.**

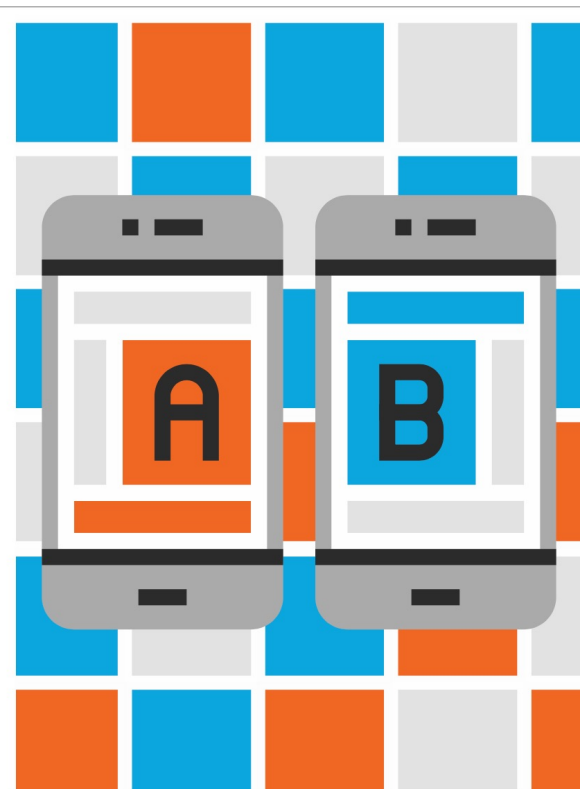
<https://www.user-testing.com/platform/userzoom>) and UserTesting (<https://www.user-testing.com/platform/userzoom>) to assist in A/B testing, enabling researchers to access real-time insights while the testing is in progress.

Although A/B testing is powerful in supporting data-driven decision making, there are drawbacks that researchers should be aware of when choosing it as a research methodology:

**The Hawthorne effect.** The Hawthorne effect describes how individuals alter their behavior or performance when they are aware of being observed [5]. While major tech companies automate their tests and users participate unconsciously, other companies often use online user testing platforms (e.g., UserTesting). Users participating in A/B testing through these platforms receive incentives and are often aware that their actions, though anonymous, are observed. This awareness might lead users to act favorably or even unconsciously modify their behavior.

**Relatively narrow and short-term focus.** The majority of online A/B testing conducted by technology companies maintains a clear goal and hypothesis. Tests are divided into atomic experiments, measuring simple user actions (e.g., the percentage proceeding to the payment page [6]). However, these tests might fail to capture whether users who didn’t proceed to payment hold a positive impression of the brand and its products. Automated A/B testing requires fully functional designs, which aren’t always feasible for various design phases. During the explorative design phase, inviting users to interact with low-fidelity prototypes can yield rich qualitative insights into the “why” behind their actions. A/B testing excels at testing focused sets of ideas, but isn’t ideal for the exploratory design phase where concrete ideas or features haven’t been defined yet.

**A THOUGHT ON AI’S TESTING**  
As mentioned earlier, leading tech companies streamline their A/B testing processes by automating repetitive tasks through scripting.



It’s important to note, however, that this automation differs from the AI technologies extensively discussed today. One key distinction between automation and AI lies in their machine-learning capabilities. Automation is rule based, following predefined instructions, while AI is trained on data, capable of learning from patterns, making predictions,

**The data collected from these ongoing tests empowers companies to make data-driven decisions.**

and even upgrading the testing scripts [6]. Recently, my colleagues and I interviewed 20 UX professionals to gain insight into their perceptions of what generative AI can and cannot do. Regarding understanding users and conducting user testing, the interviewees agreed that generative AI can help generate testing variations based on vast user behavior and preference data. It can also perform basic usability or accessibility assessments using models trained on millions of records of user data. However, it cannot replace real users participating in usability testing, nor can it validate testing results with the same level of

INTERACTIONS NOVEMBER–DECEMBER 2023 VOLUME XXX.6



Association for Computing Machinery (ACM)

Over the years, I’ve learned the importance of being proactive and realizing that seemingly insignificant tasks can lead to transformative changes.

## Memory Bites: From Earth to Space and Back

Enjoying Wine: Opportunities and Challenges for Interaction Design

Designing for Uncertain Futures

Money as an Interface

New Forum Play Time

New Column Design as Practice





What is  
Synthetic User Research?





## Synthetic User Research

It is a method that uses artificial intelligence (AI) to create virtual participants for the purpose of gathering user feedback and testing products, ideas, or concepts. These virtual participants, known as synthetic users, are designed to behave and respond in a way that mimics real users.

User research.  
Without the  
users.

User research.  
Without the  
recruitment.

User research.  
Without the  
scheduling.

User research.  
Without the  
synthesising.

User research.  
Without the  
cost.

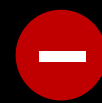




## What do experts say about synthetic user research?



Prof. Hiroshi Ishii  
MIT Media Lab



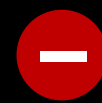
### Averaged Results

*"Generalized themes and averaged results might not accurately represent the complexity real user interactions."*

*"There is no such thing as an average user."*



Prof. Kai Kunze  
Keio Media Design



### Limited Innovation

*"Synthetic UXR that are based on established user data may provide predictions within the range of existing data but may fall short to make predictions for innovation beyond the known data range."*

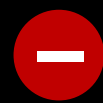


### Empathy vs. Sympathy

*"AI tools could potentially help researchers and developers maintain a balanced approach to empathy."*

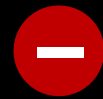


Ruben Stegbauer  
UX Manager, Google



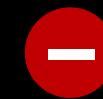
### Biases

*"...amplify our biases rather than question them."*



### First-Hand Observation

*"The contradictions between what users say and do often provide the best insights."*



### Data Validity

*"Your results will change depending on the prompt used and likely even change if you run the same prompt again."*



Dr. Sara Bouzit  
Head of Research  
EPAM France



### Under-Represented User Groups

*"AI-generated synthetic data could reveal overlooked nuances in user studies by simulating user groups that are often not represented."*



### Ethics & Confidentiality

*"We are concerned about accidentally leaking sensitive data while using AI tools such as ChatGPT for thematic analysis."*



Ye Dong  
Multilingual UXR  
Fast Retailing



### Human Touch

*"...leave out context-rich, unexpected responses linked to the personal and emotional experiences of unique users."*

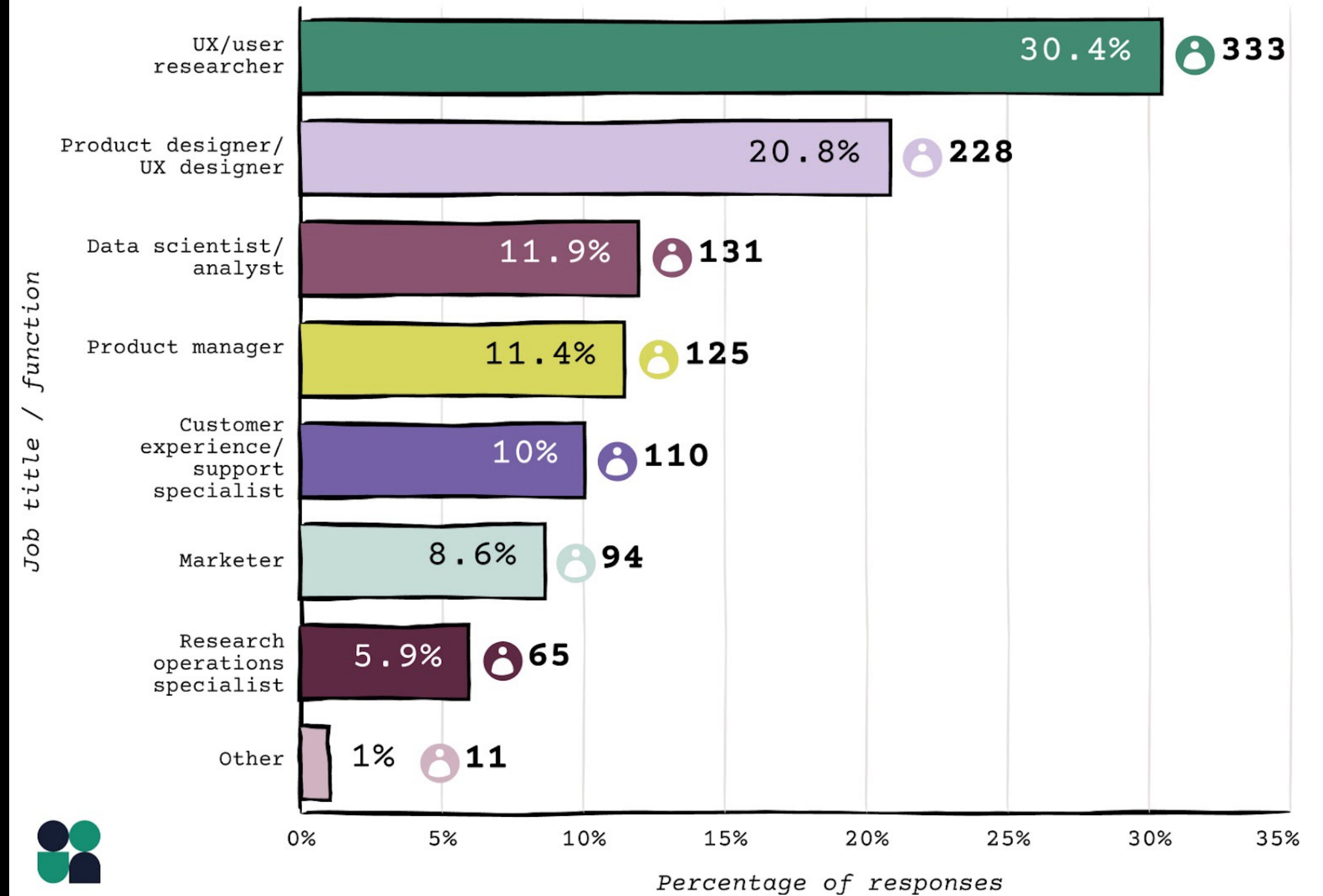
*"...lacking human touch (人味) that convey human warmth, personality, and authenticity."*



Nevertheless...

We are using AI for our work

Percentage of survey participants by job title



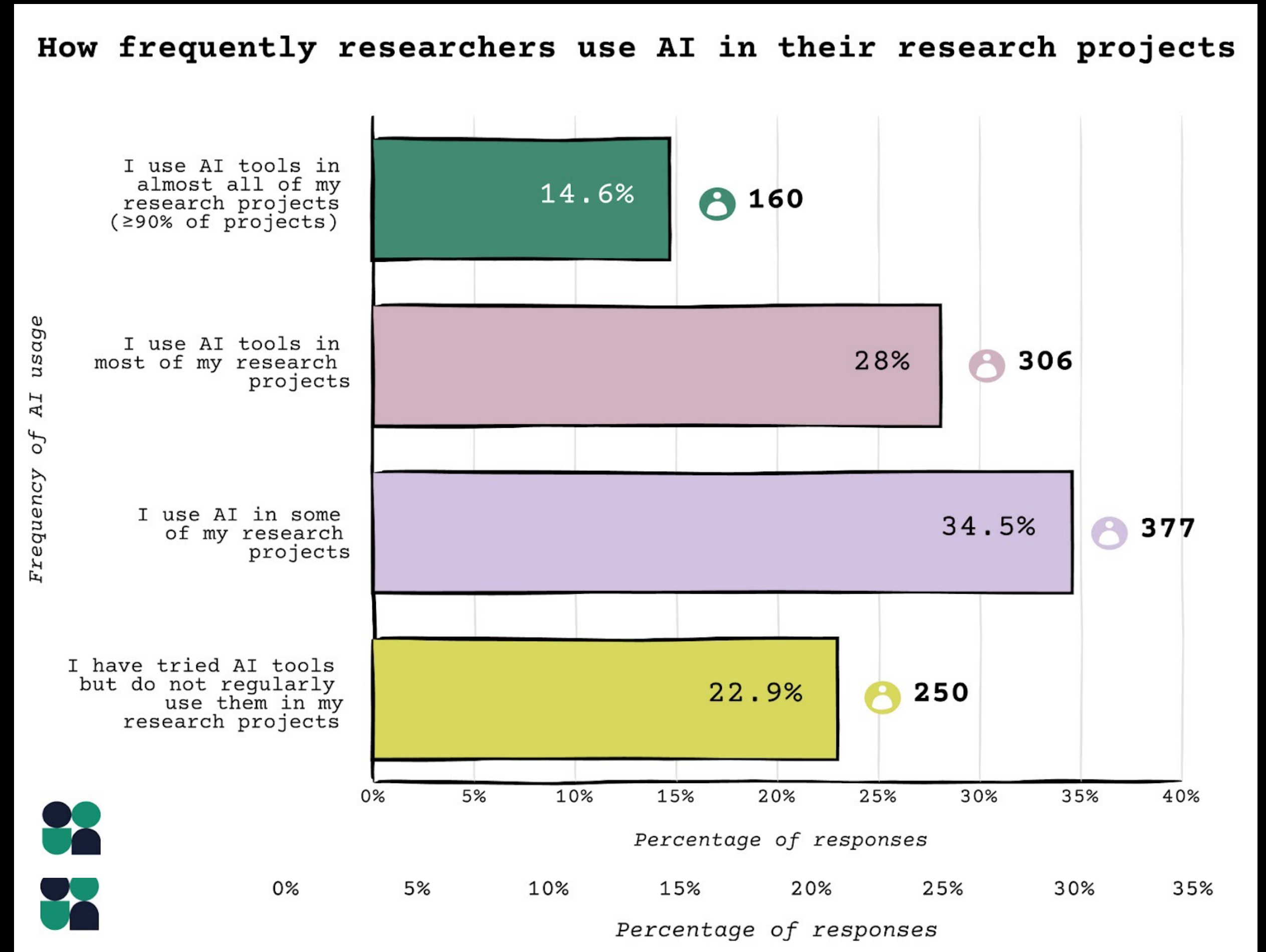
Data source: [User Interviews](#) (N=1,093)



Nevertheless...

We are using AI for our work

**77.1%** of the surveyed audience are using AI in at least *some* of their work.



Data source: [User Interviews](#) (N=1,093)

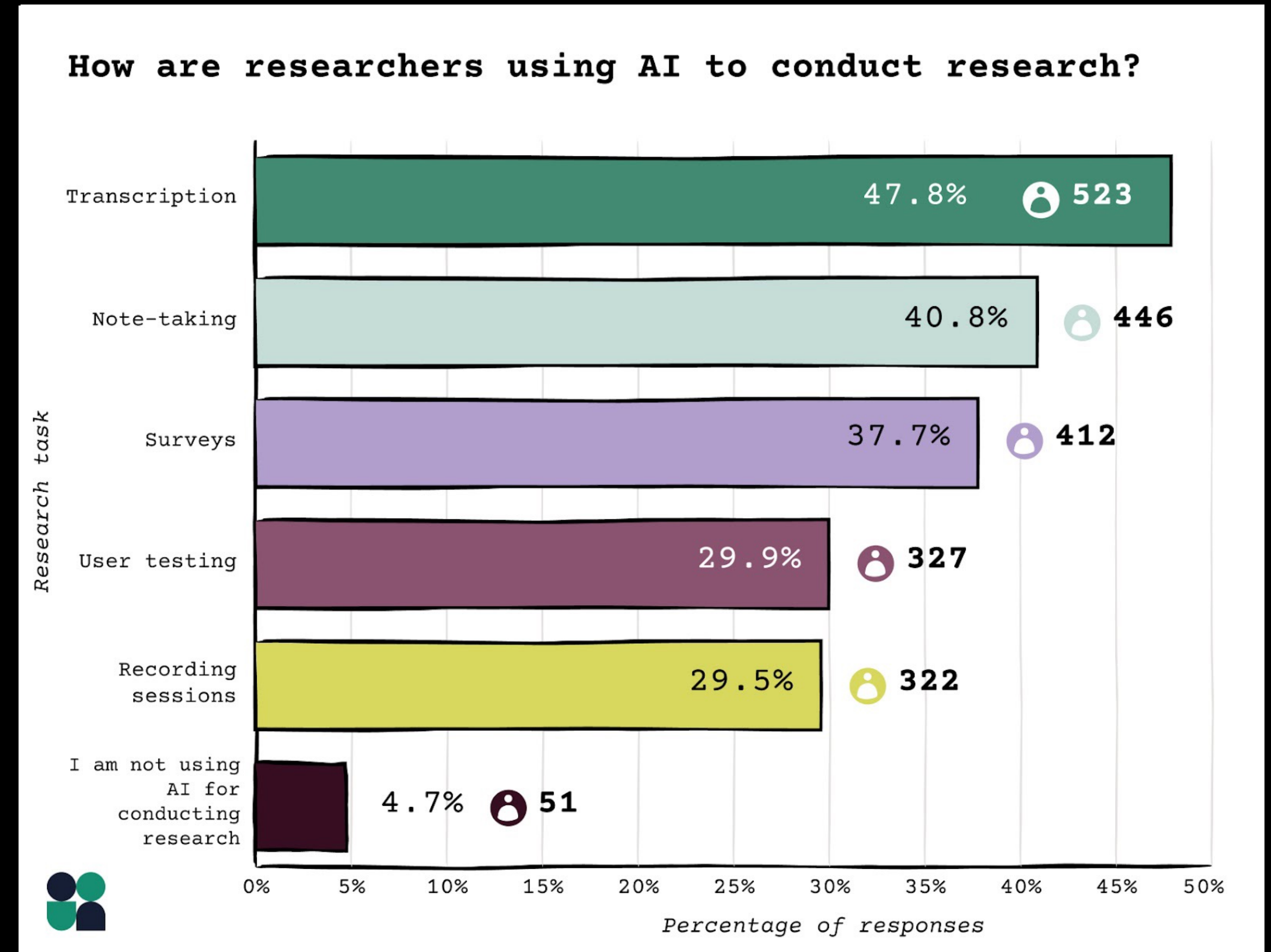


Nevertheless...

We are using AI for our work

**77.1%** of the surveyed audience are using AI in at least *some* of their work.

**47.8%** of the surveyed audience use AI for transcription and 40.8% for note-taking.



Data source: [User Interviews](#) (N=1,093)



Nevertheless...

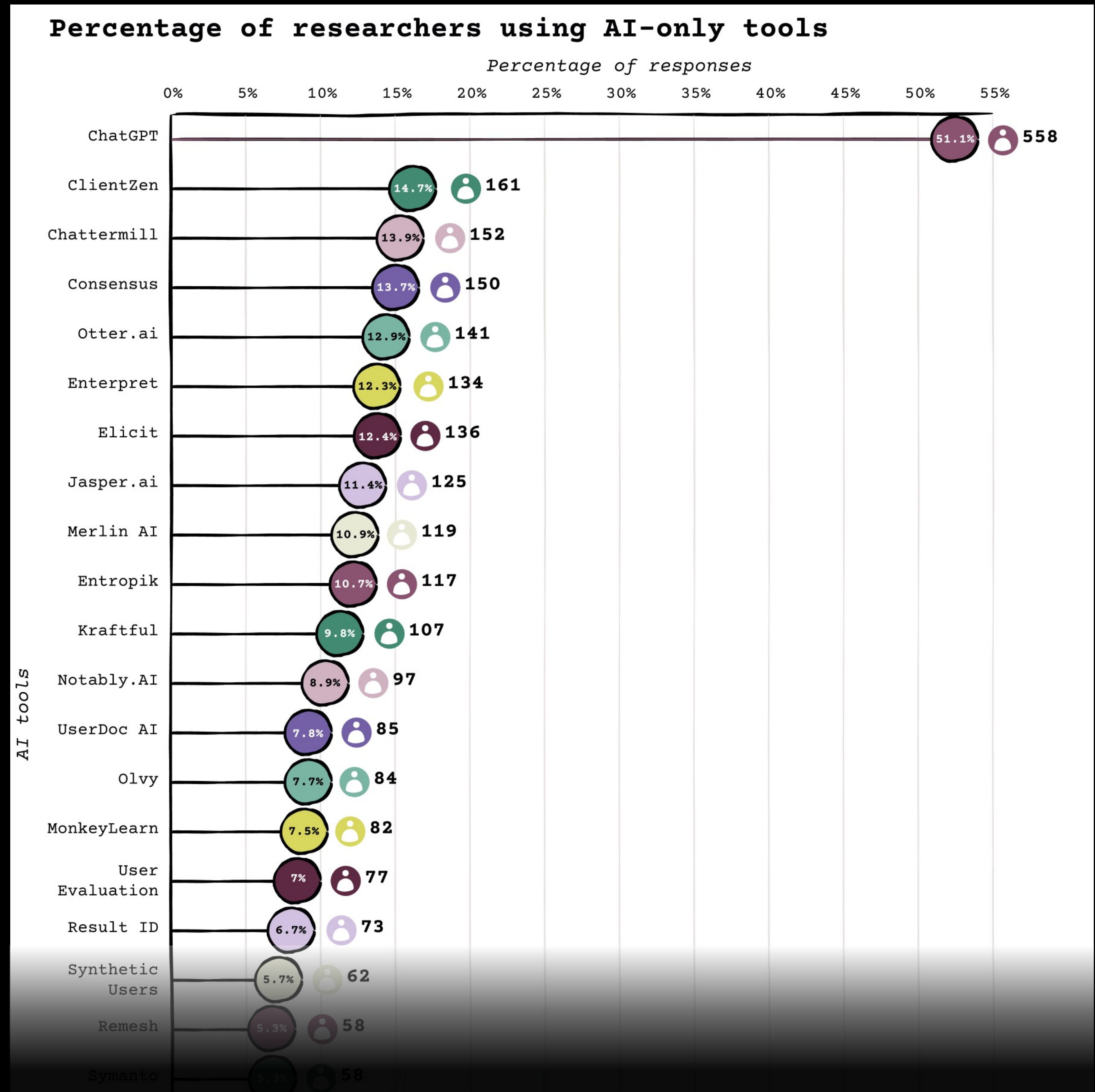
We are using AI for our work

**77.1%** of the surveyed audience are using AI in at least *some* of their work.

**47.8%** of the surveyed audience use AI for transcription and 40.8% for note-taking.

ChatGPT is the most widely used AI-specific tool, with **51.1%** of the surveyed audience use it for research.

Data source: [User Interviews \(N=1,093\)](#)





## Synthetic User Research and Code of Conduct at EPAM

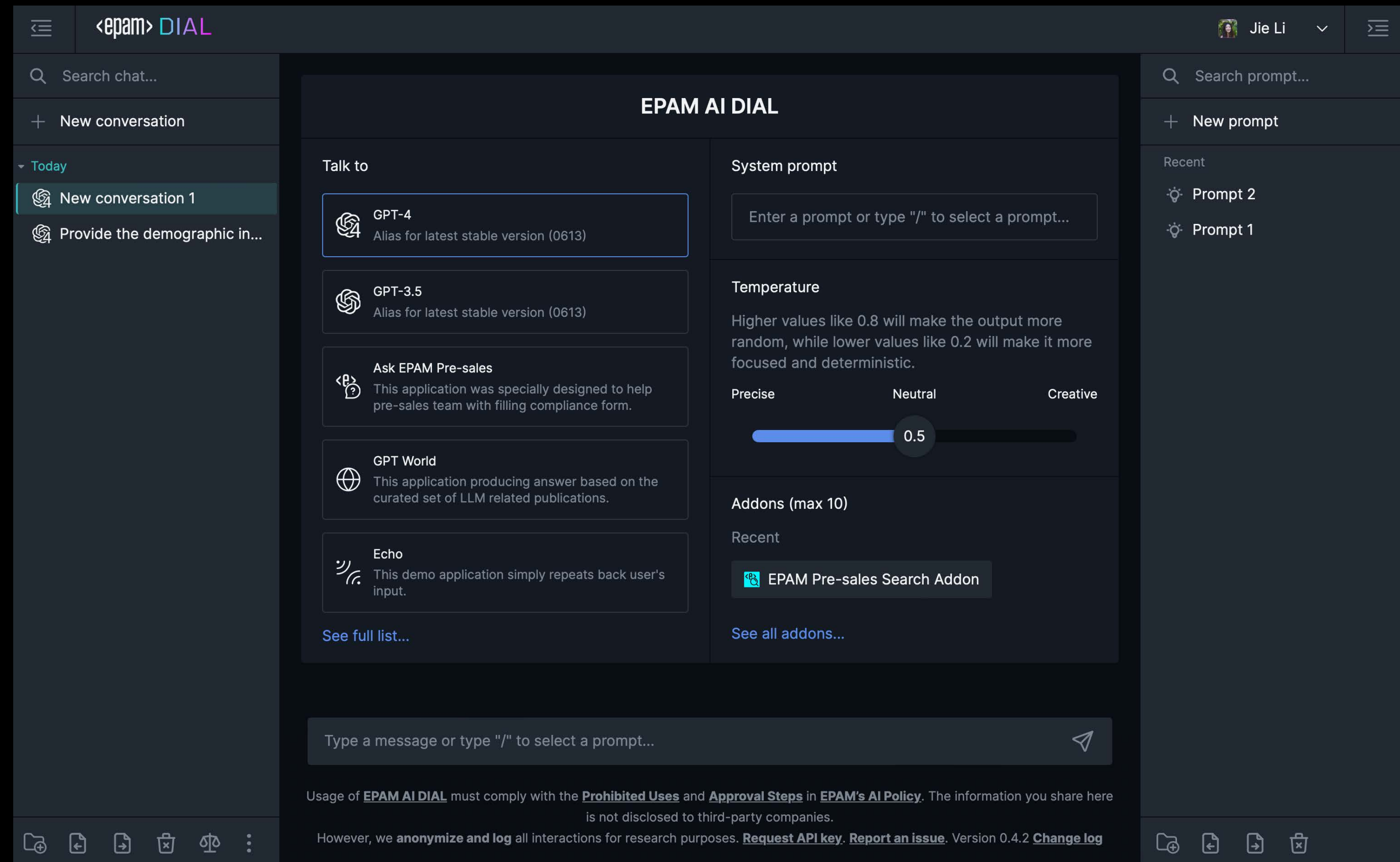
- ❌ Use GenAI tools for external or internal deliverables unless we've obtained written approval from managers and clients.





## Synthetic User Research and Code of Conduct at EPAM

- ✗ Use GenAI tools for external or internal deliverables unless we've obtained written approval from managers and clients.
- ✓ Encouraged to experiment with GenAI tools for personal tasks that are not included in any internal or external deliverables.
- ✗ Even within EPAM's in-house GenAI platform (DIAL), inputting sensitive and confidential information is not permitted.





# Some Use Cases of Synthetic User Research





## Use Case 1: Experimenting with *SyntheticUsers*

The screenshot displays the SyntheticUsers platform interface. On the left is a dark sidebar with the 'SyntheticUsers' logo and two menu items: 'Home' (with a house icon) and 'All History' (with a clock icon). The main area has a top navigation bar with four interview types: 'Dynamic Interview' (Provide the goal), 'Custom Interview' (You are in control), 'Problem Exploration Interview' (selected and highlighted in blue), and 'Solution Feedback Interview'. Below this, the 'Your study' section is divided into two columns: 'Audiences' and 'Problems'. The 'Audiences' column contains a text box describing a target audience of luxury goods enthusiasts, with a checkbox in the top right and edit/delete icons in the bottom right. The 'Problems' column contains a text box with questions about a membership program, also with a checkbox in the top right and edit/delete icons in the bottom right.

**SyntheticUsers**

Dynamic Interview  
Provide the goal

Custom Interview  
You are in control

**Problem Exploration Interview**

Solution Feedback Interview

Home

All History

### Your study

#### Audiences

☐

The target audience comprises individuals who love luxury goods and experiences and regularly purchase luxury goods or experiences and maintain memberships in several exclusive loyalty programs, spanning a diverse range of ages and genders.

#### Problems

☐

What is an ideal membership program for the target audience? What factors would encourage them to join a luxury membership program? What are the key issues and challenges associated with membership programs in the luxury industry?



## Use Case 1: Experimenting with *SyntheticUsers*

Synthetic Users

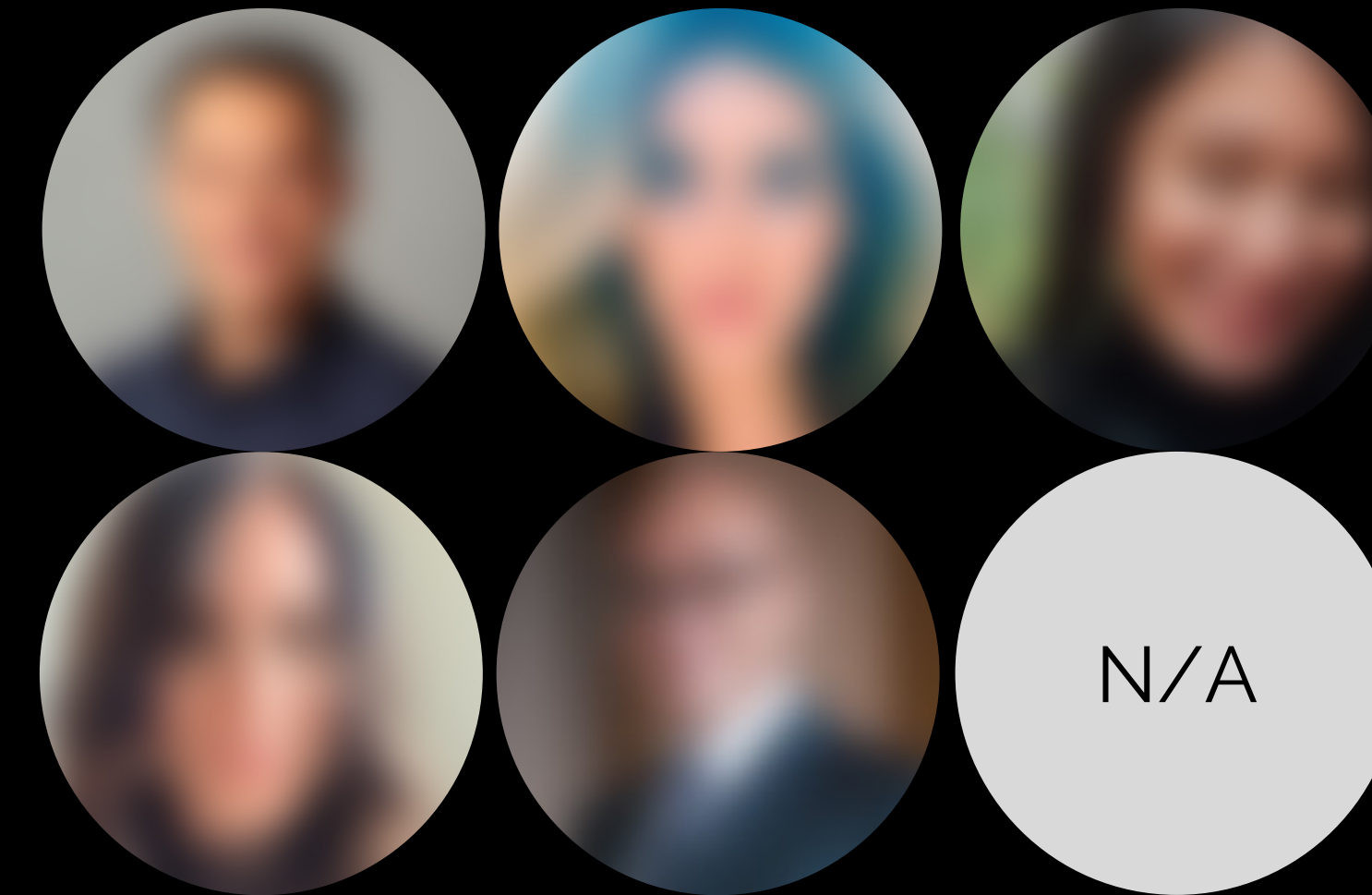


**N=6**, avoid talking about genders

**Age range:** 28-42 (M=34, SD=4.8)

**Occupations:** business tycoon, tech entrepreneur, fashion designer, investment banker, art gallery owner, fashion blogger

Real Users



**N=6**, 4 females and 2 males

**Age range:** 36-64 (M=49, SD=9.8)

**Occupations:** investment banker + company director, senior director of a tech company, international school teacher, director in a biotech company, housewife



## Use Case 1: Experimenting with Synthetic *users*

**Question:** What is an ideal luxury membership program for you?

### Synthetic Users

- Memberships that offer **unique, personalized experiences** that resonate with individual lifestyles and preferences.
- Seamless **digital integration** for easy management and access to membership benefits
- A strong focus on **sustainability**, with programs demonstrating a genuine commitment to environmental and ethical practices.
- **Inclusive exclusivity**, where programs maintain a sense of prestige while fostering a diverse and welcoming community.

### Real Users

- Emotional Bonds: **Personalized services**, recognition, and convenience
- **Digital Integration**: All services available in one place
- **Sustainability and aftercare**
- **Community**: Exclusive events within my community
- **Exclusivity**: Access to the most sought-after items
- **Quality and Longevity**



## Use Case 1: Experimenting with Synthetic *users*

**Question:** What does sustainability mean to you for luxury products and experiences?

### Synthetic Users

*"Sustainability means transparency in environmental efforts from brands I endorse and a **genuine display of concrete sustainable actions**, not just promises.."*

*"True sustainability means that the actions taken are **as luxurious and thoughtful as the products themselves**."*

### Real Users

*"My family has a tradition to use timeless products that last for years. **We are not rich enough to buy cheap stuff**."*

*"This [**aftercare**] helps reduce the disposable nature of products. If the color of my shirt fades and rest of the shirt is perfect. **I take it back and they recolor it**."*



## Use Case 2: Survey Items Validation

3. Ease to navigate/dock the boat (1 = not easy at all, 10 = extremely easy)

- This is clearly defined. However, navigating and docking could be considered separate skills and might need separate questions for a more detailed survey.

7. ~~Level of stress~~ (1 = feel stressful all the time, 10 = feel stress-free all the time)

- This item could potentially be reversed to maintain consistency with the scale direction of the other items. Typically, higher numbers are associated with more positive outcomes, so in this case, it might be more intuitive for respondents if a 10 indicated 'very stressed' and a 1 indicated 'not stressed at all'. However, this would require reversing the scale to keep the positive correlation with the other items.

**The feeling of stress-free** (1=feel stressful all the time, 10= feel stress-free all the time)



What Does GenAI Can  
and Cannot Do?





## What can GenAI contribute to research, so far?

- Aggregate information from various sources
- Perform translation and transcription tasks
- Recognize actions in videos
- Recognize emotions from video content
- Provide insights into past data and events that have occurred



Provide the demographic information of Floridian, income, household income, house price, mortgages, boat owners, type of boats, and give references for each fact



### Demographics of Floridians:

- **Median Age:** 42.6 years (Source: <https://data.census.gov/profile/Florida?g=040XX00US12>)
- **Racial Composition:** 54.9% White, 26.6% Hispanic, 14.6% Black or African American, 3.9% Asian (Source: <https://www.census.gov/programs-surveys/sis/resources/data-tools/quickfacts.html>)
- **Educational Attainment:** 87.8% have a high school diploma, 24.9% hold a bachelor's degree or higher (Source: <https://www.census.gov/quickfacts/fact/table/FL/PST120222>)

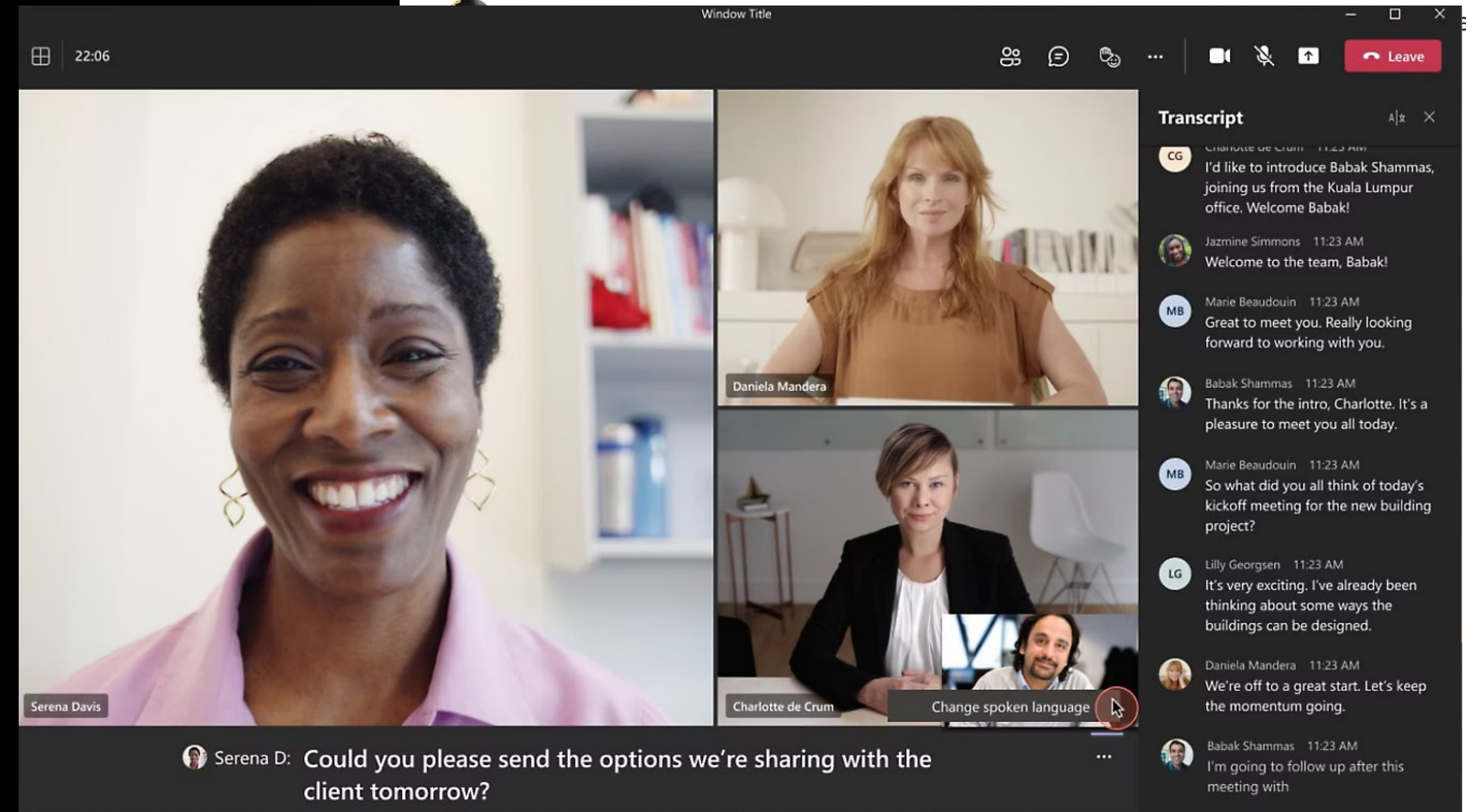
### Income and Housing:

- **Median Household Income:** \$61,915 (Source: <https://data.census.gov/all?q=Florida+Income+and+Poverty>)
- **Median Home Price:** \$470,500 (as of November 2023, Source: <https://www.floridarealtors.org/newsroom/market-data>)
- **Mortgage Rates:** Average 30-year fixed-rate mortgage - 6.09% (as of February 27, 2024, Source: <https://www.bankrate.com/mortgages/30-year-mortgage-rates/>)
- **Homeownership Rate:** 66.4% (Source: <https://www.census.gov/programs-surveys/sis/resources/data-tools/quickfacts.html>)



## What can GenAI contribute to research, so far?

- Aggregate information from various sources
- Perform translation and transcription tasks
- Recognize actions in videos
- Recognize emotions from video content
- Provide insights into past data and events that have occurred

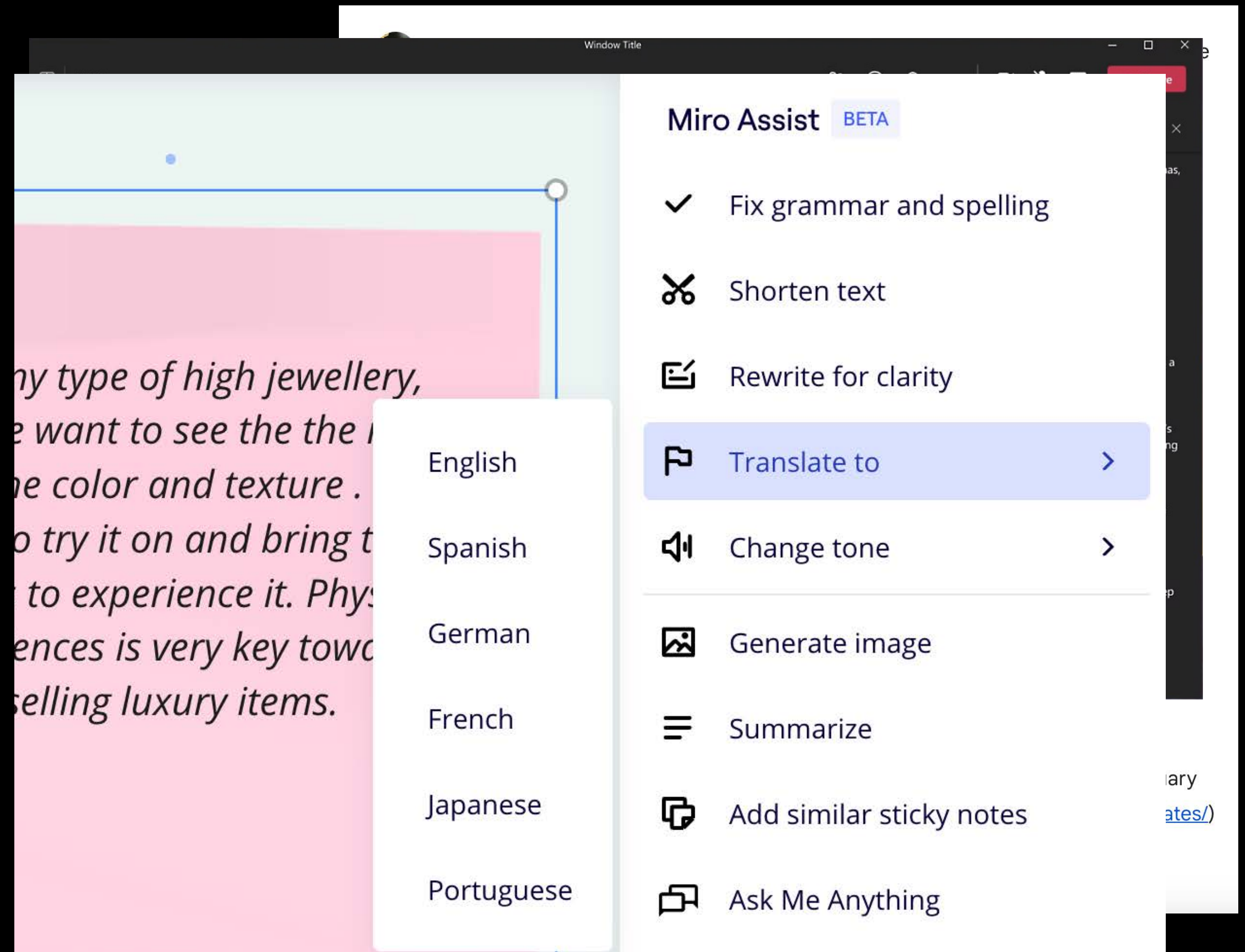


- **Median Home Price:** \$470,500 (as of November 2023, Source: <https://www.floridarealtors.org/newsroom/market-data>)
- **Mortgage Rates:** Average 30-year fixed-rate mortgage - 6.09% (as of February 27, 2024, Source: <https://www.bankrate.com/mortgages/30-year-mortgage-rates/>)
- **Homeownership Rate:** 66.4% (Source: <https://www.census.gov/programs-surveys/sis/resources/data-tools/quickfacts.html>)



## What can GenAI contribute to research, so far?

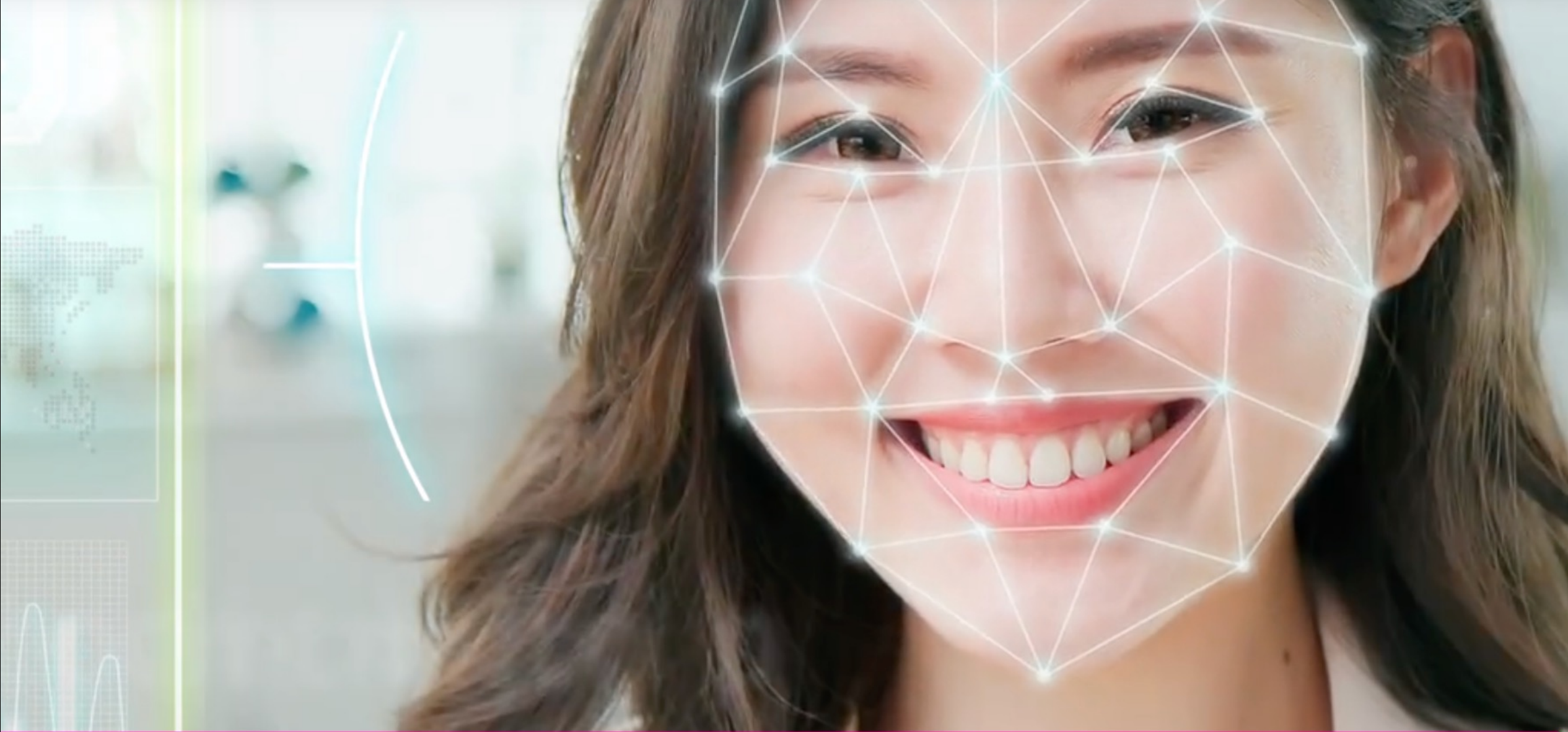
- Aggregate information from various sources
- Perform translation and transcription tasks
- Recognize actions in videos
- Recognize emotions from video content
- Provide insights into past data and events that have occurred





## What can GenAI contribute to research, so far?

- Aggregate information from various sources
- Perform translation and transcription tasks
- Recognize actions in videos
- Recognize emotions from video content
- Provide insights into past data and events that have occurred



**Afectiva Media Analytics**

Optimize content and media spend by measuring consumer emotional responses to videos, ads, movies and TV shows – unobtrusively and at scale



## AI interpolates while humans extrapolate

### AI interpolates

Generate outcomes within the range of existing data.  
Excel at repetitive tasks within known data patterns.

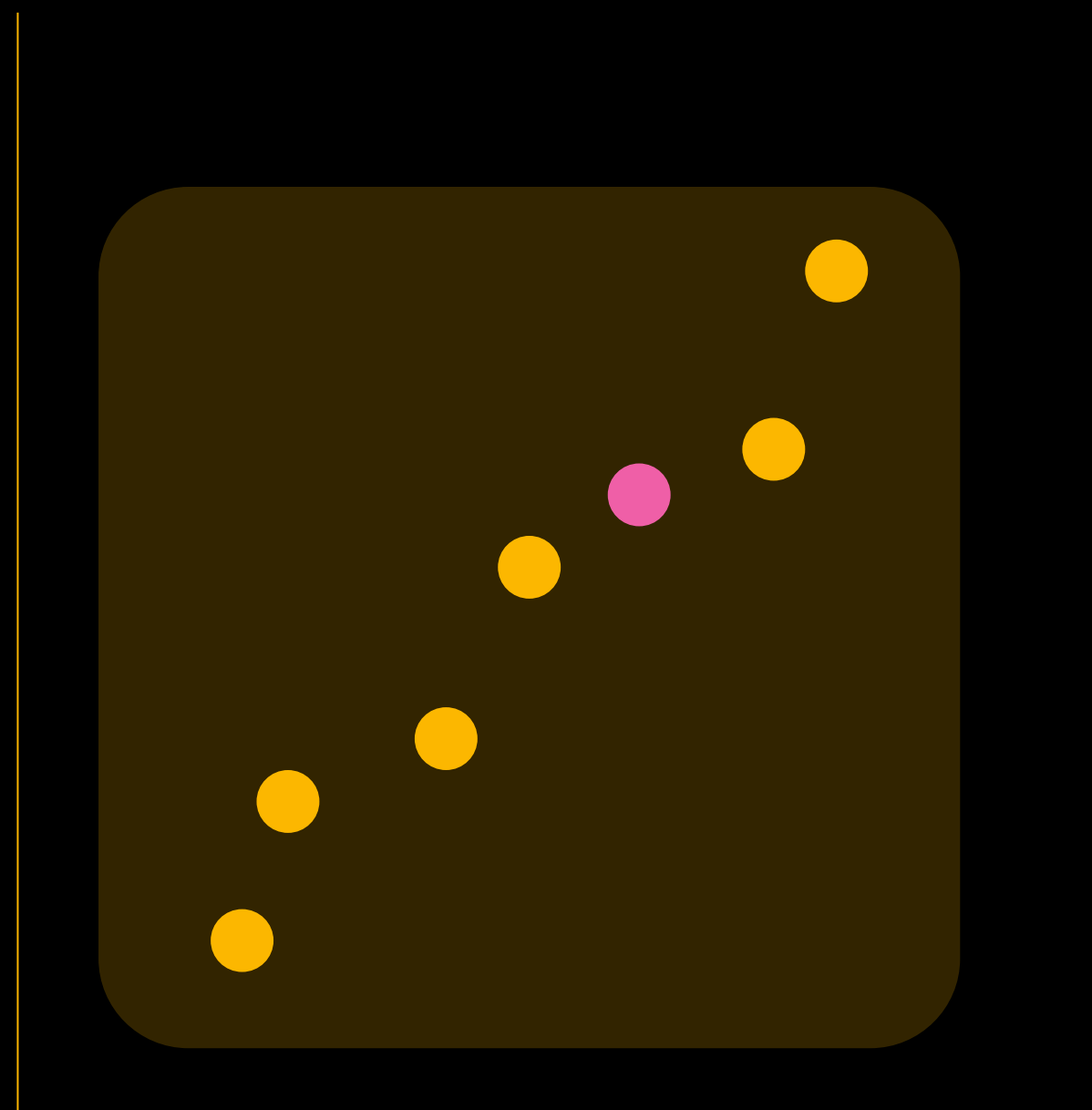
### Humans extrapolate

Make predictions beyond the known data range, using our creativity, intuition, empathy, domain knowledge, sensory experiences, social interactions...

### Children are good examples

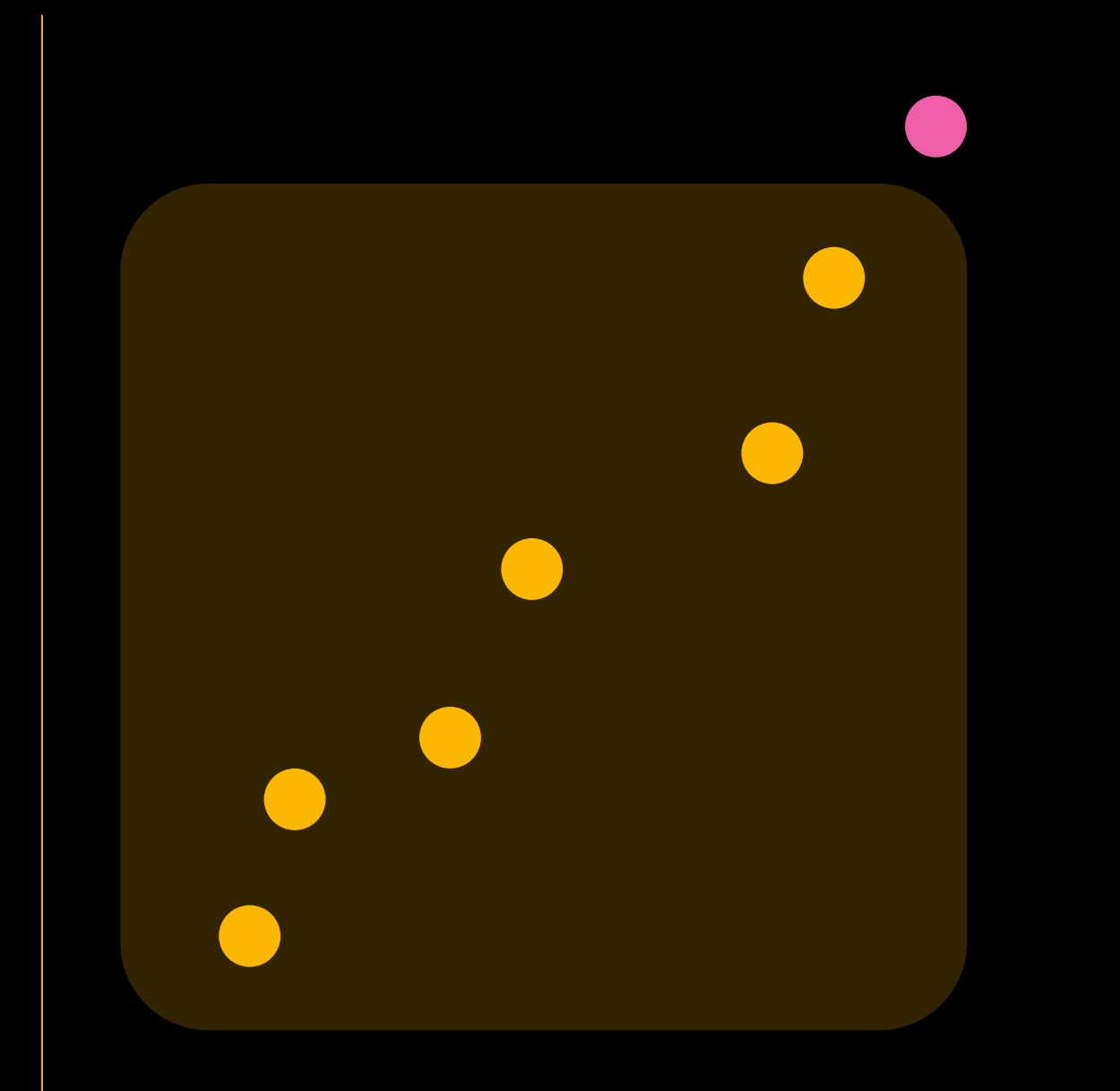
### Interpolation

Predicting within known data range



### Extrapolation

Predicting beyond known data range





## AI interpolates while humans extrapolate

### AI interpolates

Generate outcomes within the range of existing data.  
Excel at repetitive tasks within known data patterns.

### Humans extrapolate

Make predictions beyond the known data range, using our creativity, intuition, empathy, domain knowledge, sensory experiences, social interactions...

### Children are good examples



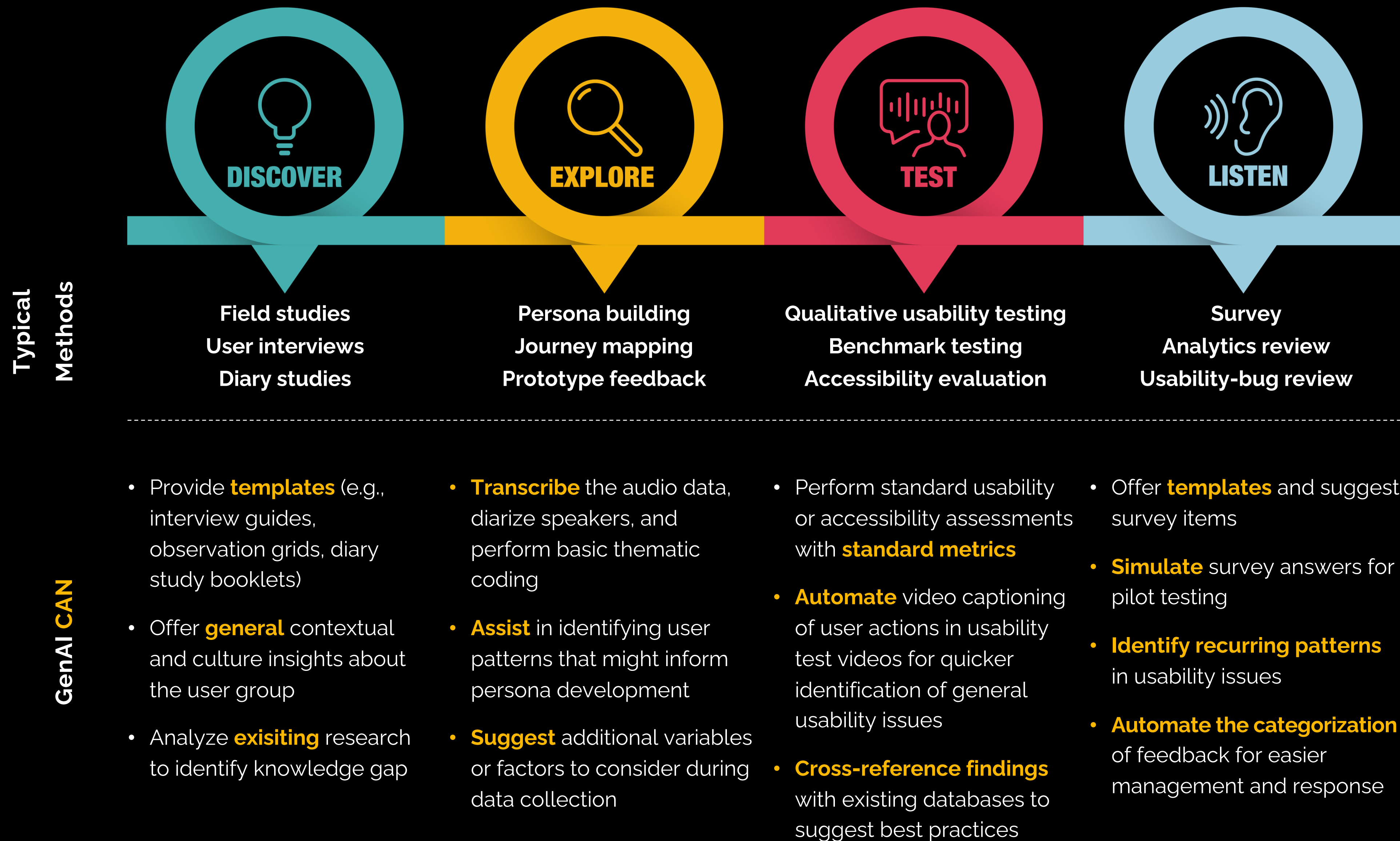
An illustration generated  
by DALLÉ-3

The illustration captures the essence of children learning languages and showcases their ability to exceed the capabilities of large language models through diverse methods of language acquisition, creativity, and the joy of learning.



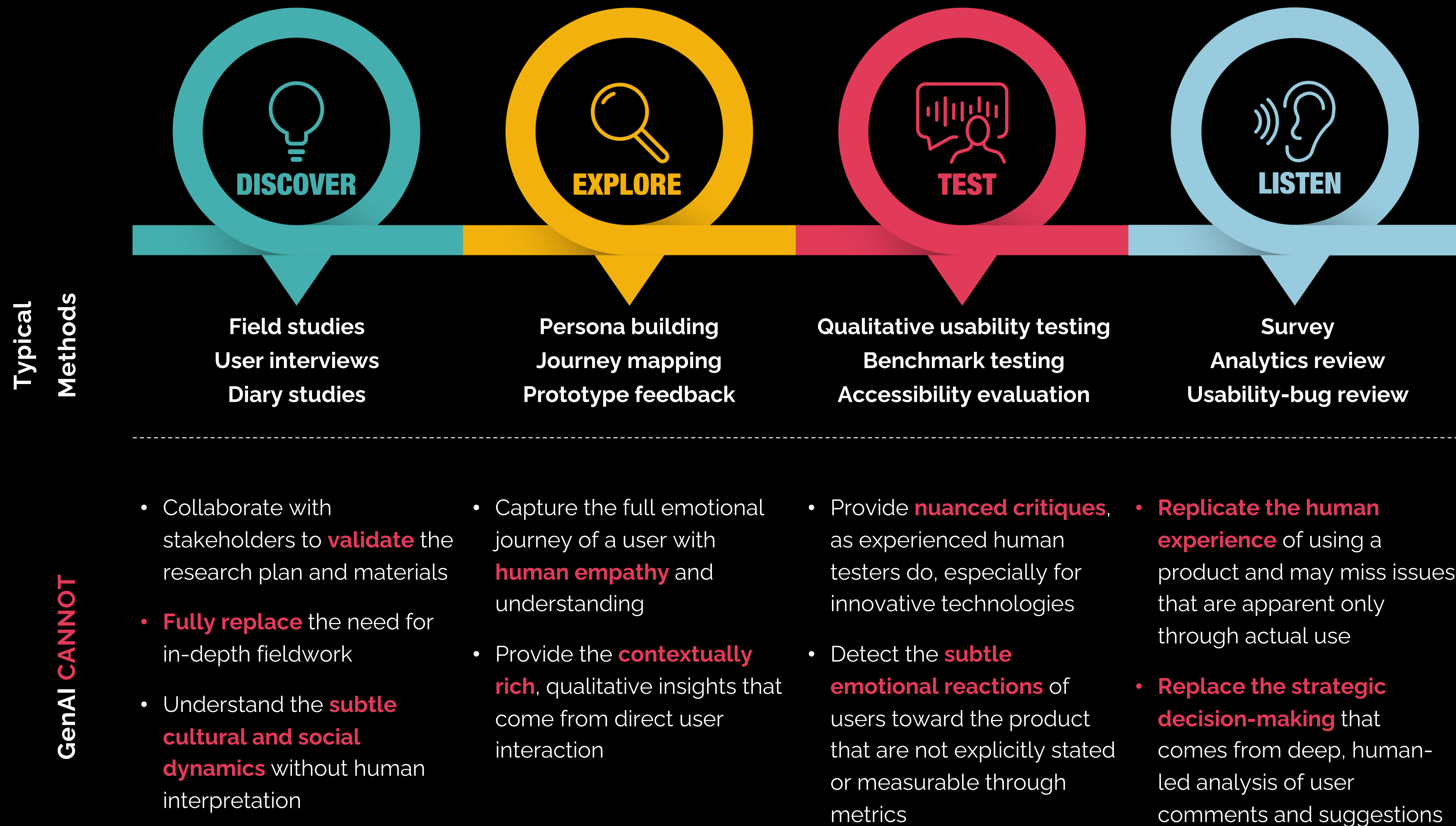






The four stages of a UX research cycle: Discover, Explore, Test, and Listen (Farrell, 2017)









## QUESTIONS?



[jie\\_li@epam.com](mailto:jie_li@epam.com)



[www.linkedin.com/in/jie-li-nl](https://www.linkedin.com/in/jie-li-nl)



[jieli-research.com](http://jieli-research.com)



[@cake\\_researcher](https://www.instagram.com/cake_researcher)