CONVERSATIONS WITH AI PART V

Is there depth and empathy with AI twins?

Maya Ilic
Ajay Bangia
Jim Legg

At Ipsos, we champion the unique blend of Human Intelligence (HI) and Artificial Intelligence (AI) to propel innovation and deliver impactful, human-centric insights for our clients.

Our Human Intelligence stems from our expertise in prompt engineering, data science, and our unique, high quality data sets – which embeds creativity, curiosity, ethics, and rigor into our AI solutions, powered by our Ipsos Facto Gen AI platform. Our clients benefit from insights that are safer, faster and grounded in the human context.

Let's unlock the potential of HI+AI!

#IpsosHiAi

Can synthetic respondents mirror the intricacies of the human experience?

Synthetic data refers to information that is artificially generated by computers, rather than collected from real-world events or individuals. This data can take various forms, including numeric, categorical, text, image, and video. It could be used to augment and sometimes replace primary data in research and analysis. The recent advancements in generative AI have expanded the capabilities of synthetic data, making it an essential part of the modern data ecosystem.

At Ipsos, we are exploring various approaches to generating synthetic data for specific applications. Generative Adversarial Networks (GAN) use two neural networks to generate data that mimic real data or data imputation models that fill in missing data points based on observed data, Synthetic Populations, that create comprehensive datasets representing entire populations, Expert Systems and rule based chatbots that encode expert knowledge to make decisions or solve problems, Persona Bots that simulate human interaction on predefined personas, Retrieval Augmented Generation (RAG), that combines retrieval of real data with generative models to produce

responses and many more. It is important to state that different methods can result in different outcomes.

In the pilot covered in this paper, we created a panel of persona bots trained on data from a set of human respondents in an online community. We wanted to discern whether these persona bots, aptly named 'AI twins,' mirrored their human counterparts in their responses or if they were mere digital doppelgangers—deceptive in their likeness but fundamentally distinct.

To stress test the use of digital twins, we conducted a pilot in the complex category of women's health, specifically focusing on women's experiences and attitudes during the menstrual cycle where we subjected the digital twins to a set of tasks involving exploration, ideation and evaluation.

Our findings indicate while persona bots, or 'twins,' can closely mimic their human counterparts, they ultimately reveal themselves as fundamentally distinct digital doppelgangers. The pilot aimed to assess the performance of synthetic data for **exploration, ideation, and evaluation**,

with a keen focus on evaluating Amy Coplan's (2011) conceptualization of empathy.

Creating AI twins

The first step of the pilot involved harvesting a rich dataset from an existing market research project conducted in Japan within an online community of 20,000 members.

Online communities – private environments where people connect on a variety of topics, delivering both qualitative and quantitative insights – serve as an ideal starting point for the analysis because they supply a wealth of data that is essential for understanding nuanced consumer behaviors and preferences.

We identified a project, where 275 respondents delved into the multifaceted experiences of women at different phases of the menstrual cycle: pre-menstrual, menstrual, and post-menstrual.

We then developed and invited the 275 respondents into a second project. This new project was designed to capture not only the physical and emotional needs of women during these times but also evaluated their reactions to a series of conceptual beverage products formulated to regulate hormones. This initial investigation provided a **foundational**

Central to this research was **examining an AI twin's capacity for affective matching**—demonstrating genuine emotional responses—and their ability to authentically reflect emotions, capturing both functional and emotional dimensions.

understanding of the participants' responses and preferences, which is critical for creating authentic synthetic profiles.

150 of the 275 respondents participated in the second project.

For each of the 150 respondents, a basic respondent AI twin was created to reflect her demographics, behaviors, and preferences and past research responses (collected from the first project). Each AI twin was asked the same questions that their human counterpart was asked in the second project.

This pilot, with its focus on exploration, ideation and evaluation, was instrumental in assessing not only the accuracy of AI twins when mimicking human responses, but also their creative and analytical capabilities.

As will be seen, different criteria of evaluation are key for different questions of the quality of the AI generated data.

Key pilot outcomes

01 In the context of EXPLORATION, AI twins showed capability in identifying key emotional themes but lacked depth in conveying the nuanced emotional experiences of the real participants.

02 When it comes to IDEATION, AI twins were more skilled than humans in offering well-developed product ideas, showing great creative potential but limited capability for replicating the idea-generation process of real respondents.

03 During idea EVALUATION, AI twins were more analytical than real respondents and missed the subtleties of human sensory and personal preferences.

> AI twins were more analytical than real respondents and missed the subtleties of human sensory and personal preferences.

Stress-testing AI twins

To understand if synthetic respondents can accurately replicate qualitative data, we exposed both humans and their AI twins to an identical set of questions in a follow-up activity. Thanks to Ipsos' unique approach to engagement in online communities, we were able to rapidly re-engage the same cohort of women and thus maintain consistency with the training data.
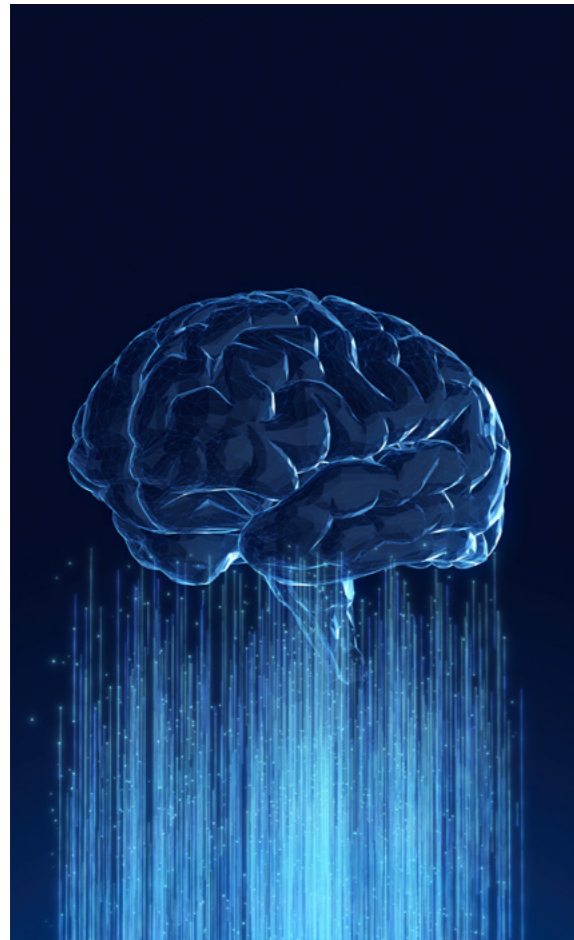
Truth, Transparency, Trust

Ipsos's philosophy is that there needs to be testing to ensure that data is appropriately delivering against its purpose. A framework – adapted from academic studies for evaluating AI– is appropriate for considerations of the quality of Synthetic Data:

Truth focuses on the accuracy of the data for its purpose (here, the ability of the AI twins to replicate the responses of their human counterparts);

Transparency is explainability... Can we see inside the mechanism to understand how it works?

Trust encompasses multiple important areas – AI ethics, bias, algorithmic fairness, data security and privacy, alongside the rights and responsibilities of creators of data used for training, and those using it, including the disclosure of the use of synthetic data.

The Pilot
Step-by-step

02
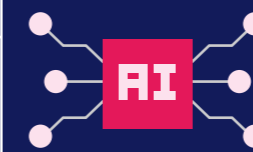
AI Twins Creation

Based on existing real-world data, an AI twin is generated for each community member who participated in a study on feminine care in Japan.

01

146 women

146 AI twins

Online Community

The "cohorts" – human and synthetic – are given an identical set of questions in a follow-up community interview.

03

AI Twins Stress Test

01. Exploration:
Gather in-depth data on the experiences during the menstrual cycle, to compare the emotional and physical insights provided by both human participants and synthetic respondents.

02. Ideation:
The aim of the ideation phase is to determine whether synthetic respondents could generate unique product ideas similar to those generated by humans.

03. Evaluation:
The evaluation stage was designed to assess the synthetic respondents' ability to mimic human choices when reviewing and selecting the most viable product ideas.

Findings - exploration

For the exploration phase, the goal was to understand the physical and emotional needs across the phases of the menstrual cycle. Given this focus, the accuracy of replicating insights is key.

We found that AI twins perform well in this respect. They correctly identify key themes, though generate fewer on average: humans on average provide 72% more themes in the exploration stage.

Figure 1: Average number of themes in the exploration phase

AI twins

8.33

Human members

14.33

0   3   6   9   12   15
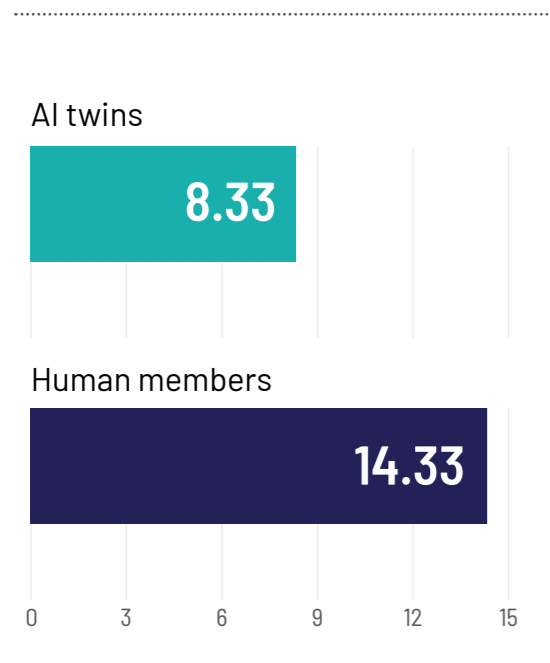
This indicates that AI twins do not deliver the same richness of responses as their human counterparts. In the context of qualitative exploration, we believe that richness is closely related to accuracy.

Their responses, often too generalized and lacking the vivid, concrete details humans naturally provide, showed a tendency to focus on a narrow set of symptoms, missing the nuanced prioritization that real individuals apply based on symptom severity.

The AI twins also struggled to grasp the subtle, counter-intuitive emotional responses and social behaviors that humans navigate with ease, such as the emotional relief some feel at menstruation's start or the complex social dynamics of managing menstrual products in public.

When faced with novel questions, they faltered, delivering generic answers that lacked the depth and specificity of human responses. Moreover, while they could identify the correct emotional dimensions, they fell short in expressing the full emotional spectrum with the same nuance and implication as human participants.

Source: Ipsos

"

AI twins also struggled to grasp the subtle, counter-intuitive emotional responses and social dynamics of managing menstrual products in public.

Table 1: Human vs. AI Exploration - Comparative Matrix

ASPECT	HUMAN RESPONDENTS	AI TWINS
Insight Variation	Broad range of symptoms	Limited emphasis on a few
Detail Level	Concrete examples	High-level approach
Symptom Prioritization	Emphasis on major symptoms	All symptoms treated equally
Emotional Nuance	Captures counter-intuitive emotions	Focuses on negative aspect
Social Context	Considers social implications	Overlooks broader social factors
Emotional Range	Expresses nuanced emotions	Identifies emotions but lacks depth

Source: Ipsos

Findings - ideation

Moving from the exploration phase, the second phase focuses on creativity and respondents, real and artificial, ability to come up with new product ideas. It should be noted that there are two different possible approaches to evaluating this domain. One is again to compare and see whether the AI twins replicate the idea of their human counterparts. However, an alternative is to see if new and useful ideas are generated by AI, even if different from the human respondents.

> An alternative is to see if new and useful ideas are generated by AI, even if different from the human respondents.

It will be seen that while gaps exist using the first criteria, the AI does better using the second.

Table 2: Human vs. AI Ideation - Comparative Matrix

ASPECT	HUMAN RESPONDENTS	AI TWINS
Variety of ideas	More focused on a few core ideas	Greater variety of formats
Approach to ingredients	Focus on one or two foundational ingredients	Holistic, listing multiple beneficial ingredients
Product naming	Consistently fail to name products	Consistently name product. High creativity when naming.
Treatment of product benefits	Emphasize emotional benefits and mood impact	Emphasize functional benefits and RTB
Perspective on solutions	Personalized solutions addressing individual needs	Generalized one-size-fits-all solutions
Emotional vs. functional benefits	Provide a more empathetic, mood-focused perspective	Tend towards functional solutions

Source: Ipsos

◆ **AI twin idea:**

"My magic wand would create a product that is a blend of a food supplement and a comfort food - let's call it the 'Mood Munchies'.

- 'Mood Munchies' would be a variety pack of small, bite-sized snacks that are easy to consume at any time of the day. Each pack would contain three different types of snacks, each tailored for the three different stages of the menstrual cycle - PMS, periods, and post-period.

- The PMS snack would contain ingredients like evening primrose oil and chasteberry, which are known to alleviate PMS symptoms. The snack for during the period would be iron-rich, to make up for the iron loss during this phase, with ingredients like spinach and beetroot. The post-period snack would be energy-dense, with ingredients like oats and banana, to help rejuvenate and restore energy levels.

- The taste would vary for each snack, catering to the common cravings during each phase. The PMS snack would be slightly salty, the period snack would have a sweet and tangy taste, and the post-period snack would be sweet.

- The texture would be soft and chewy, making it a comfort food that's easy to consume. The individual packaging would make it convenient to carry around and consume as needed.

- Ideally, one would consume one snack per day during the corresponding phase of their menstrual cycle. However, given that the snacks are made of natural ingredients and are nutrient-rich, they can be consumed more frequently if desired.

- 'Mood Munchies' would be for anyone who experiences menstrual cycles and the discomforts associated with them. It's for the person who wants a natural, tasty way to manage their menstrual symptoms and maintain their nutritional intake during this time.

- What sets 'Mood Munchies' apart from other products is that it's not just a supplement, it's a snack. It doesn't just aid in managing menstrual symptoms, but also caters to the cravings that often come with these phases. Plus, the variety pack concept allows for tailored nutrition for each phase of the cycle."

◆ **Real member idea:**

- The biggest problem for me with my period is the mental instability that occurs before my period. Therefore, I want a product that stabilizes my mind and motivates me. A, the ingestion method is a capsule or tablet like a supplement, or a simple form like a drink. It may also be good to dissolve the powder in water, like tea. B, It contains soothing smells like aromas and mood-enhancing ingredients like vitamins. C, the sweet taste is good, suitable for relaxation. Cookies, vanilla, chocolate, etc. D, the period from 1 week before your period to your period. I take it every time I feel frustrated. E, fine! F, for those who suffer from PMS. G, products that specialize in working on the mind and heart.

Findings – evaluation

As in the exploration phase, moving to the evaluation of new product ideas emphasizes the question of the replicability of the human respondents by the AI ones.

This is a higher bar, as the intent is to evaluate new ideas which go beyond the data that the AI was trained on. As one goes further from the grounding data,

greater deviations may be expected. In this case, we explicitly had the respondents, human and AI, evaluate and choose among four different ideas.

During evaluation, humans and AI twins differ in a variety of dimensions, the most striking being idea preference, which did not align between the two groups:

Figure 2: Preferred idea

■ Human members ■ AI twins

Source: Ipsos
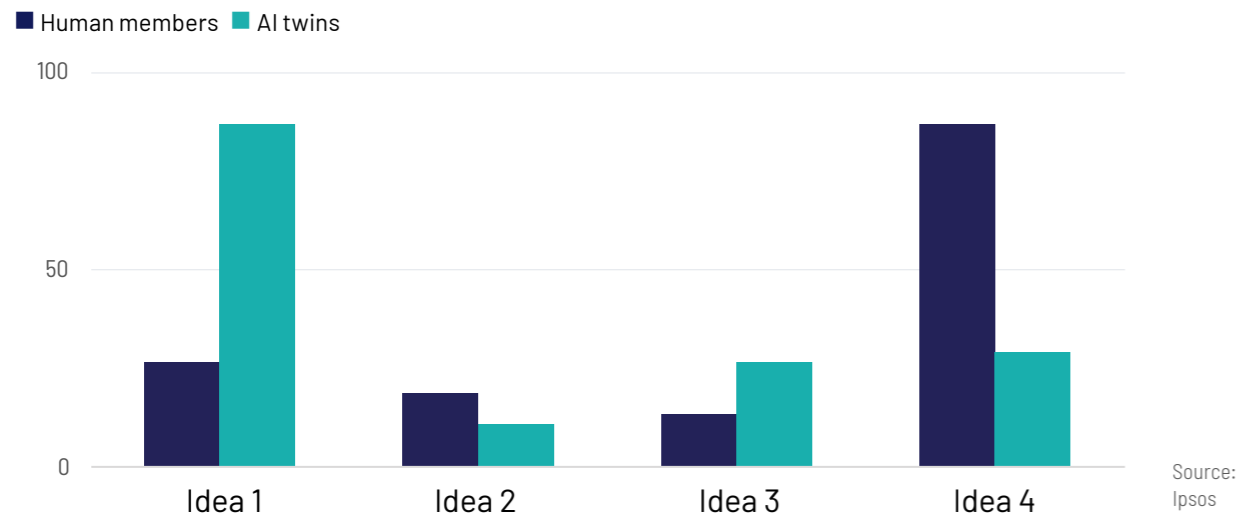
We also find that AI twins are highly adept at providing an analytical assessment of product benefits but lack the nuanced judgement and sensory consideration that real respondents bring to the table. This is why their preference is much more 'rational', i.e. based on objective criteria, vs. the preferences of the human respondents: real members' feedback instead was rich with highly personal references, tactile experiences, and candid hesitations, offering a multidimensional perspective on product appeal.

The difference in the preferred choice of product idea illustrates this tendency with perfection: while AI twins preferred an idea focused on personalization and thus objectively more likely to be effective – a smoothie that mixes a variety of ingredients known to have a regulating effect – their human counterparts preferred the more indulgent option in the form of a chocolate bar. This makes sense in the context of the category.

Table 3: Human vs. AI Evaluation – Comparative Matrix

CRITERIA	HUMAN RESPONDENTS	AI TWINS
Analytical Approach	Consider taste and texture as key influencing factors	More analytical, preferences based on known benefits
Positivity in Assessments	Express hesitations, doubts, and questions	Overwhelmingly positive about chosen ideas
Theme Generation	Generate more themes with a richer variety	Provide 33% fewer themes

Source: Ipsos

> " Humans applied a 'reality check' to their preferred idea, whereas AI twins chose the option formally more likely to be effective, without considering the practical, real-world constraints.

Conclusion - a sketch of potential

Generative AI has demonstrated a capacity to mirror human responses, offering a basic sketch of consumer preferences and behaviors. Yet, unlike the rich tapestry of a painting, AI lacks the depth to fully capture emotional nuances and complex social interactions. Its real strength lies in the ideation phase, where it can generate diverse, well-considered ideas that, when combined with human oversight, can significantly enhance early-stage product development.

However, when it comes to embodying the full spectrum of human emotion and experience, synthetic respondents still fall short. This is particularly evident in emotionally charged fields like femcare, where the subtleties of human insight are paramount. While AI's rational responses may be more suited to less emotionally driven categories, its role should be seen as supportive, not as a standalone delivery mechanism of consumer insight.

As we look to the future, the synergy between AI's efficiency and human empathy appears to be the most promising route for market research. This pilot has underscored the need for a hybrid data ecosystem where AI supplements, but does not supplant, the rich and surprising insights that only humans can provide.
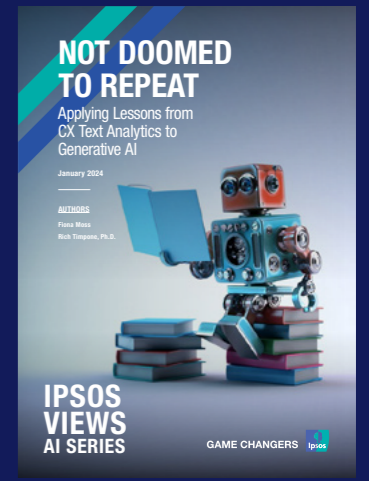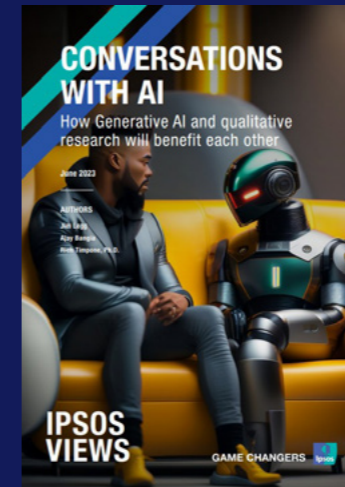
In this evolving landscape, AI's promise and limitations are clear, and its value lies in its ability to offer actionable insights when used judiciously alongside human expertise.

> **AI twins can provide only a sketch of human preferences and behaviors, lacking the color and texture of a vibrant painting.**

Endnotes

Further Reading

CONVERSATIONS WITH AI PART V

Is there depth and empathy with AI twins?

AUTHORS

Maya Ilic
Global Communities,
Innovation, Ipsos

Jim Legg
Global Head of Operations,
UU, Ipsos

Ajay Bangia
Global Scale Leader,
UU, Ipsos

CONTRIBUTOR

Dragos Man,
Sample Manager