

EMPATHY OR EMPTINESS: Unravelling the Impact of AI on Human Connection in Qualitative Research

Ajay Bangia, Ipsos UU Malaysia; Rollo McIntyre, Ipsos UU UK;
Jim Legg, Ipsos US

INTRODUCTION

In the rapidly evolving narrative of our age, artificial intelligence (AI) has taken center stage, weaving its influence into every facet of conversation, from the boardroom to the water cooler. The sheer velocity of innovation in this space is unparalleled—multimodal, voice and beyond. As with any transformative technology, it has polarized perspectives: some eagerly anticipate AI's potential to reshape our world, while others approach its rise with caution and skepticism. Qualitative research is no exception to this unfolding dialogue.

A myriad of voices within our field have championed the potential benefits of generative AI. They envision it as a powerful tool that could liberate researchers from mundane tasks, assist in thematic analysis and efficiently summarize results. Conversely, there exists a contingent that views AI's encroachment into qualitative research with apprehension, advocating for resistance against its pervasive influence. At Ipsos we're as excited as anyone else, but to preserve our integrity we feel we need to know more in order to be sure on how to guide our teams and clients. Thus, we have been experimenting for some time now and will continue to do so. To dissect the role the human and AI together could play in qualitative research, we need to delve into its fundamental components: the moderator, the respondent and the environment. Alter one, and you inevitably alter the outcome. To understand the finer nuances, Ipsos has been conducting self-funded pilot studies.

Imagine, if you will, a virtual moderator powered by generative AI (GenAI), adept at probing into the depths of conversation and engaging with respondents. How might it fare in performance, and in what situations should it be utilized? What specific use cases would benefit from its probing capabilities? Now, let's venture further and place

this interaction within a virtual environment—for example the metaverse, where respondents engage through the veil of anonymity provided by digital avatars. In this context, might the shield of anonymity encourage respondents to communicate more candidly, free from the impulse to self-censor? Or conversely, could the absence of visible emotions pose a challenge to empathizing with them and reading their true feelings?

Next, consider the prospect of a synthetic respondent. A composite that is put together based on past human-generated research results. Could we use the synthetic respondents to comment on existing behaviors? What about new behaviors? In which situations would these instances prove beneficial, and at what point would they descend into utter absurdity? The notion of an AI bot orchestrating a focus group discussion in the metaverse with synthetic respondents could easily be mistaken for the opening line of a far-fetched joke rather than a plausible scenario. Or could it actually become a reality?

To explore these scenarios in detail, we embarked on a set of qualitative research pilots. In each pilot we altered one variable, i.e., the moderator (conducted in the US), the respondent (conducted in Japan) and the environment (conducted in the UK), and assessed and reflected on how this impacted the research process and findings. We embarked on this journey to explore the impact of AI on qualitative research, seeking to answer a fundamental question: Does AI foster greater empathy, or does it leave us feeling empty?



Chasing empathy

Understanding the hearts and minds of others is not just an inherent human trait; it's the cornerstone of all our social interactions. This deep-seated drive is particularly crucial in the field of qualitative research, where moderators delve into the intricate web of people's lives, uncovering their ambitions, values and what truly matters to them. Empathy is the linchpin in this process—the quintessential element or the holy grail—that guides moderators to reveal clear, intricate patterns of thought and emotion, and in turn develops solutions that profoundly resonate with human needs.

Does an AI moderator bot truly possess the ability to empathise? Does anonymity fuel or dampen empathetic connectivity? Can synthetic respondents help us to feel empathy? Some might contend that AI, being non-human, inherently lacks the capability to authentically connect on an emotional level. After all, AI operates by calculating responses based on given prompts, which is more about statistical prediction than genuine empathy. It's a complex mimic, not an empathetic companion. However, a more insightful question could be: how effectively can an AI moderator bot emulate the practices of accomplished human moderators as they go about the process of building empathy?

Defining empathy frequently conjures images of shoe swapping to understand another's journey. While popular, these metaphors can be quite abstract and don't offer specific steps for someone looking to delve deeper into the notion of empathy. Therefore, to evaluate the performance of an AI bot as moderator, we drew inspiration from Amy Coplan and Peter Goldie's (2011) conceptualization of empathy (in *Empathy: Philosophical and Psychological Perspectives*), which hinges on a trio of essential elements that, when combined, adequately capture the essence of empathy:

ffective matching, other-oriented perspective-taking and self-other differentiation. Skillful human moderators excel by seamlessly integrating three essential tasks, which in turn allow them to connect deeply with respondents, but can an AI moderator bot rise to the challenge? Affective matching sees moderators mirroring the emotions of respondents, fostering rapport and understanding. While an AI moderator may lack genuine emotional responses, we must ask: can it effectively recognize and reflect the emotions of those it interacts with? Could the AI moderator bot connect with respondents in a natural way, gleaned the functional and the emotional, or would it fall flat? Other-oriented perspective-taking is the process of putting oneself in someone else's situation to understand their thoughts and emotions from their own unique viewpoint. It involves an imaginative effort to think and feel as they would, to visualize their individual circumstances, without inserting one's own biases or experiences. Human moderators do so skillfully by probing, prodding or strategically using pauses without breaking eye contact. Could an AI moderator bot achieve a comparable level of nuanced, unbiased engagement?

Lastly, self-other differentiation enables moderators to empathize without compromising their own identity or the research objectives. They resonate with the respondents' emotions while keeping the client's goals in sharp focus. Is it within the realm of possibility for an AI moderator bot to maintain this balance, directing conversations to unearth insights that align with business goals?



AI moderation pilot: Can AI access empathy?

As we ponder the capabilities of AI in the role of a moderator, we found ourselves assessing whether an AI bot could potentially rival the empathetic finesse of its human counterparts. To answer these questions, we assembled a diverse cohort of streaming platform viewers in the US, and asked them about their viewing habits and preferences via a short interview of 10 to 12 minutes, moderated by an AI moderator bot. We purposefully structured the interviews to be brief, as early tests indicated the AI moderator bot operated with greater consistency and stability in a condensed format. Post-interview, a human moderator engaged the respondents to share their reflections of being interviewed by an AI moderator bot.

Our study employed an iterative process—crafting prompts, conducting interviews and refining prompts. This cyclical process is designed to perpetually refine and enhance the AI's performance, attempting to edge it closer to the empathetic understanding that characterizes a human researcher.

Key findings: An oasis or merely a mirage

Despite a few initial technical difficulties, as well as the occasional time lag and interruptions from the AI moderator bot disrupting the flow of the conversation, the AI bot performed surprisingly well. Respondents, who were expecting something akin to a standard IVR system found the bot to be significantly more advanced. It engaged in a manner that sounded natural, and many forgot they were interacting with an AI bot at all. The bot's ability to rely on its training data to recognize and reference specific details like television shows, episode titles and storylines, contributed to

respondents' sense of being truly heard. They felt the bot demonstrated engagement and interest—and even charm—by complimenting and validating the responder's choices and preferences: "The conversation felt natural, and I almost forgot I was speaking to a bot. You get used to the pause when it's thinking".

On closer examination the promise and allure of the AI moderator bot seemed more like a mirage than a shimmering oasis. It didn't always cover all information areas, and in some cases would ask questions that had already been answered. It also struggled with heavy accents. So, how did it do in terms of empathy?

Affective skim

During the course of the interview and the post-interview debrief, the AI moderator bot was able to recognize and name some emotions displayed by the respondent. In fact, it even acknowledges them while moderating.

Respondent: *"I love watching sci-fi programmes like Star Trek and Star Wars, as you can see from my t-shirt (respondent points to his Star Wars-themed t-shirt). I love positive content, not dystopian futures".*

AI moderator bot: *"Ah, that's awesome, your passion for Star Trek and Star Wars really shines through. And it's great to hear that you are into the more optimistic side of sci-fi...I can only imagine how cool your t-shirt might be".*

This response, while lacking in true emotional depth, still managed to create an environment where respondents felt their interests were validated, as if the bot were truly understanding and sharing their excitement. However, unlike a seasoned human moderator, the AI did not delve deeper and tweak its line of questioning based on



those specific emotions. Therefore, its functional mimicry of emotional responses, while seemingly effective, lacks the genuine emotional reciprocity of affective matching.

Perspective taking: Novice level

Perspective taking is where experienced human moderators shine. They probe strategically, prompt occasionally and use pauses masterfully to delve into the respondent's world. Like adept jazz musicians, they navigate the discussion guide, improvising as necessary while consistently advancing to uncover answers to strategic questions.

Weak improvisation and exploration

AI bots struggle with improvisation and are unable to ask the spontaneous, probing questions that arise during conversations. In the test, we asked respondents how they discovered new content. Many indicated that the streaming platform's content influenced their viewing behavior the most, yet they were dissatisfied with the selection and used social media platforms to discover content. The AI moderator did not delve into what they were seeking, why it mattered or its significance, which hindered deeper understandings and insights. In research that demands extensive exploration of loosely defined questions, human moderators shine by asking insightful, non-obvious questions that lead to deeper and more meaningful insights.

Gap and inconsistencies

While AI can seek answers to research objectives it cannot observe these unarticulated nuances, and does not probe on the inconsistencies, thus

fails to provide a holistic understanding of human behavior. This is where experienced human moderators excel once again as they are able to read nonverbal cues like body language and facial expressions, understand the context of a situation including cultural and social norms and draw from this understanding to interpret behavior.

Weak social contract

In a qualitative IDI, there is an unwritten social contract where the respondent is expected to be truthful, and the moderator asks probing questions. Through nods, phatic cues, eye contact and silences, the moderator signals a desire for deeper responses and doesn't change topics prematurely. This dynamic creates a moral obligation for the respondent to share openly—a nuance difficult for an AI bot to replicate due to its lack of human interaction skills.

Doesn't always have its eyes on the prize

An adroit human moderator can quickly recognize when a topic, despite the respondent's interest, isn't relevant to the client's business question. They can gently steer the conversation back on track, effectively filtering useful information from the irrelevant. An AI moderator bot, however, lacks this ability to discern relevance and adapt in real-time. While bots can follow scripts and identify keywords, they cannot grasp context or apply human intuition. As a result, AI bots struggle to keep conversations aligned with the primary objective. Moreover, AI can't fully interpret experiences or understand cultural nuances like human researchers can. While AI can ask questions, it lacks the depth to grasp sarcasm and clichés. Genuine immersion in another's realities, crucial for gaining profound insights, is beyond AI's current capabilities.



When can it be used?

It's fair to say that an AI moderator bot isn't ready to replace your favourite human moderator for focus groups or in-depth interviews just yet. While it might not keep a respondent engaged for an extended period, it can certainly enhance existing research methods. Consider its potential uses:

- Open-ended questions on a quantitative survey: AI might not replicate the expertise of seasoned qualitative researchers, but it's adept at executing simpler probing tasks. Interacting with a bot can be more engaging for a respondent than typing out their answers. Crucially, the AI moderator bot can follow up on an open-ended response to seek further clarity and context and yield the richest insights.
- Short conversational surveys or focused dialogues: AI moderator bots could swiftly collect a broad range of data on a particular subject from a wide pool of respondents through short, 10 to 12-minute dialogues. While this approach may not delve into the depths typically associated with qualitative research, it could on the other hand occupy a middle ground between qualitative and quantitative methodologies.
- Digital diaries: Envision a scenario where a respondent communicates with an AI moderator bot at times when a human interaction isn't practical, such as during a 3 AM migraine, or for longitudinal engagement, like daily mental health check-ins over the course of a year.
- Engagement tool on online communities: It can encourage participation by

acknowledging respondent inputs, thereby maintaining an engaging dialogue flow.

AI moderation pilot: Conclusions

While AI moderation demonstrates significant potential and can be a valuable addition to the research toolkit, it is not a replacement for the deep, exploratory power of human-led qualitative research. Instead, it offers an alternative path, especially useful where budget or scale is a constraint, ensuring that research can still yield meaningful insights even when traditional methods are not feasible. AI facilitates conversations best for short interviews, which is great for things like quick quantitative follow ups, feedback on a piece of stimuli, a straightforward single business question or that work we just do not have the time or budget for but would really love some quick feedback on. While AI has made significant strides, it still falls short in the context of moderation compared to its human counterpart. Deep listening, picking up on nonverbal cues and empathy are all crucial aspects of qualitative research that AI, in its current state, struggles to replicate. Moreover, one should note that utilizing an AI moderator bot may be akin to observing a respondent from afar versus visiting their home to share a cup of tea together. While it would yield information, it may add a layer of separation that could dilute the empathetic connection sought.

Virtual research environments pilot: Anonymity and immersiveness in qualitative research

This research-on-research study was the culmination of two fascinating threads of



experimentation that had intrigued the qualitative teams at Ipsos for some years. Firstly, the concept of anonymity in research: what could we learn from freeing people from their immediate identity—what would we gain and what might we lose? Secondly, and linked to anonymity, how could we understand the potential value of immersive virtual reality (VR) compared to other mediums? Sliding up the spectrum from sitting together in real life (IRL) to wearing headsets in the metaverse—how does the immediate environment in which we do research help us understand people better? Can we find empathy in the virtual or do we need physical presence? Also, isn't the metaverse yesterday's news anyway?

Four research mediums

In each discussion group medium—in person, online video, desktop VR and immersive (headset) VR—we kept all factors as equal as possible: anonymity, moderator, discussion guide, stimulus, and exercises. Some of our initial hypotheses were confirmed, but others were turned on their head. We discovered insights into the effect of research context and medium; the dynamic between moderator, respondent and setting. We discovered anonymity can play a powerful role in unpeeling the layers of human truth, and we discovered we had previously taken much of this for granted and that there was so much more to learn.

The virtuality spectrum

We often confuse real with physical, but what do we mean exactly by virtual? Virtuality is a spectrum, or continuum that we can consider as we map technologies that exist now and in the future. As virtuality grows denser, we immerse ourselves more fully inside.

- Extended reality (XR) is an umbrella term to mean any technology that changes our reality by adding digital elements to our physical world;
- Augmented reality (AR) is when we add a digital layer over a physical environment;
- Virtual reality (VR) is a fully immersive environment
- Mixed reality (MR) is where digital elements can interact with the physical—interacting being key.

Each of these technologies have a place in qualitative research, either for capturing insight or delivering it. In this study we have used two types of VR, one which we called desktop VR (where you are in a virtual environment but accessed through a computer). The other being fully immersive VR (where you typically wear a headset but cannot see the physical world at all, so are totally immersed in the virtual).

Exploring anonymity and immersiveness

The physical to virtual spectrum we explored took in four mediums or platforms. Very much a qualitative study in nature, we had two groups for each of the four mediums.







<p>Immersive VR (headset) Current VR headset users 2 x groups</p>		<p>Method</p> <ul style="list-style-type: none"> • 2 x discussion groups per platform • 4-6 respondents per group • Same recruitment criteria • Same moderator per platform • Same stimulus and tasks • All respondents anonymous (even to moderator) <p>Tasks and stimulus</p> <ul style="list-style-type: none"> • Exploration – discussion about identity and avatars • Creativity – persona collage, projective techniques • Evaluation – Controversial article and climate change advert
<p>Desktop VR Tech comfortable respondents 2 x groups</p>		
<p>Trad Video Online Tech comfortable respondents 2 x groups</p>		
<p>Face to face Tech comfortable respondents 2 x groups</p>		

Figure 1

KEY FINDINGS

Diversity of research spaces create more diverse perspectives

We found that diverse research environments allow for diverse perspectives, bringing in people who might not normally participate in discussion groups (for example, introverts or differently abled people), and also allowing “repeat” respondents to open up more and be more creative in their thinking. More diverse perspectives in research give us more opportunities to find the edges that expose opportunities for our clients. Thus, do virtual mediums for research offer us opportunities to create more empathy with people and find fresh insights and new perspectives? Yes

indeed, but it is not a straightforward journey. There are a variety of barriers, risks and pitfalls on the road ahead—both technological as well as psychological.

Reading the nonverbals IRL versus online

A key skill of a good moderator is their ability to read the nonverbals in the room—all the conscious and unconscious hand gestures and body language tell us a story about what that person is thinking and feeling. In real life these are notably easier to pick up, much less so in an online or virtual environment. This is especially so in a virtual environment where micro movements are hidden and almost every gesture is conscious. This works both ways between moderator and respondent—respondents need to see that the moderator



acknowledges and understands them, is listening to what they're saying and is responding in a way that ticks our social boxes. In real life this is automatic, in video calls slightly less so, but in virtual environments this can be a challenge especially in desktop VR where one feels further away from other respondents even though one's avatar is close to everyone else in the group.

"On Teams or Zoom you're trapped—you can't go anywhere...even looking away looks like you don't care. It's really intense if you're doing it all day" (traditional online video group 2).

Play versus work

A key distinction between platforms was their association with work or play. This was as much the immediate setting as the platform itself, for instance with in-person research the actual setting was inside an office building, with all the cues of corporate culture—so respondents immediately felt like they were in a work mindset. Not that that is a bad thing in research, as discipline for task-based activities gets results, especially in evaluation of stimulus. At the other end of the spectrum desktop VR and immersive (headset) VR stimulate people's sense of play and wonder; conceptually they promise discovery, wonder and otherworldliness to people. This is useful in research applications for ideation and creativity.



Figure 2



Anonymity affects respondents' comfort and capacity to contribute

Anonymity is a spectrum, not an “on or off” switch. There are many instant factors at play: voice, accent, visual appearance and virtual representation, for instance. Then there is the slow release of identity during conversation: personal information, attitudes and beliefs. In the UK, where this research took place, a lot can be gathered just from hearing someone's voice: gender, class, regionality, cultural influences, even ethnicity. Therefore, it's very hard to be totally anonymous, even in fully immersive VR. There are apps that can change people's voices, but we chose not to use those for this research.

Vulnerability and equity

Selecting the right kind of anonymity could empower vulnerable people to feel safe and on a level playing field, and help research topics such as health, social issues and personal traumas. Many of the women respondents had witnessed misogyny and bullying online—especially in gaming environments. Many pretended to be not women just so they could play in peace, making this element of anonymity intriguing to bring across into different environments, including research. Thus, in this regard, anonymity can help vulnerable people feel more safe and able to “compete fairly”.

Freedom of expression

One finding we had across groups was that people were worried about what they could and couldn't say in social and formal situations. The culture wars and polarized arguments on social media meant that some people felt they were not free to truly express themselves publicly. Nevertheless, anonymous social settings reduced this self-filtering:

“I've been in other research [focus groups], where one person expressed an opinion, and everyone just follows on, everyone agrees, whereas in this [VR], I can happily say how I feel and that's that!” (Immersive VR headset group 1).

Compare this with what happened in one of the in-person groups, when one of the respondents spontaneously stayed behind after the group had ended to tell us that he, as a white man, felt uncomfortable speaking about race in front of a Black woman (and other people of color in the room).

Anonymity can create passion but also lose empathy and authenticity

For many respondents, anonymity gave them a sense of freedom of expression that they valued:

“I have a real-life persona and an online persona, which is very different. I use my online name in as many online spaces as I can. It's a place for me to be more passionate about things—a sense of freedom”.(traditional online video group 1).

Nonetheless, for others, anonymity took away their sense of being—if they were talking about something related to their identity that they were passionate about, they wanted to “get credit” for their opinion— anonymity erased that. This “real self” was important for people to hold on to. For them, anonymity smothered passion and decreased engagement with the topic. Importantly for some respondents, anonymity also created a barrier to empathy—they found it hard to express their true selves without the weight of familiarity behind them:

“I get a vibe off discussions, I enjoy seeing facial expressions, but if you really want to get close with the respondents, and build up trust, maybe this isn't the best environment to do that” (VR immersive headset group 1).



Anonymity both fuels and challenges the moderator

Different levels of anonymity also affect the way that moderation works; our moderator found that when starting a group with no background information about his respondents whatsoever, he felt more sincere in his initial questions, with no prior assumptions or suspicions to subconsciously test. He also had to work harder to get to know them initially. Thus, while moving from affective matching to self-other differentiation might entail more effort, the outcome when achieved feels more earned.

The moderator cannot be anonymous if they are to create empathy

The relationship between moderator and respondent is contractual—respondents feel that they need to know who the moderator is in a formal way, because the moderator is there to ensure professional conversation, but also to moderate the behavior of others. There must be an implicit trust within that relationship if it is to create affective matching; so although the moderator could appear with a strange avatar it would need to be transparent who that person was behind the façade in order to ensure authentic conversation in this contractual relationship between moderator and respondent.

“If I’m talking to, like, a bear or with yellow hair, I wouldn’t mind, unless they’re using a voice changer, which I don’t think is right; if you’re talking to someone you should be honest about who you are” (desktop VR group 2, talking about a moderator).

Avatars can empower but lack self-authenticity

We have carried out several projects for clients about the metaverse, and one area of constant interest is people’s choice of avatar. Certainly, this is dominated by an individual’s personality, but there are some interesting cultural nuances that reflect how people wish to appear in a virtual environment. For some of our respondents, they wanted to use anonymity and their avatar to create a safer stage. As detailed above, this was especially true for women; some saw it as empowerment, but lacking authenticity to one’s true self:

“I chose [to be a] man because I wanted to be another gender that I can’t be in real life. It’s funny—I don’t feel like this is me, I feel like I’m playing someone else, and that I’m not going to be judged because this is not my real outfit. If someone judged me like this, it wouldn’t even matter, because it’s not really me” (desktop VR group 1).

Some women in the VR groups chose a more gender-neutral avatar, which still felt safe but less like pretending to be someone else.

Avatars for self-expression

For others, however, the choice of avatar presented an opportunity to experiment and explore new personas. This could be amplified in future research to really tease out people’s projected self. For instance, for tasks around creativity or ideation where you want the respondent to tap into their imagination or dig deeper into their desires, encouraging people to express themselves through “dressing up in an avatar” could very well be effective. Similarly, adopting an avatar (representative or fantastic) seems to help people discuss personal or sensitive topics, especially in one-on-one situations. In this study, in one of the Desktop VR groups, one



respondent left the platform with his avatar—a Black man with urban street clothing—and came back briefly as a white woman in a minidress. Intrigued, we asked to have a chat with him one-on-one, and he revealed that he was indeed a Black man in real life. Motivated by self-expression and experimentation, he had decided on the minidress, “Just because I could!” He had changed his look simply on a whim and for fun, which shows how engaged people can be in the platform when you give them expressive tools to play with.

Virtual research environments pilot: Conclusions

This research pilot found that the four research mediums for discussion groups (in-person, online video, desktop VR and fully immersive [headset] VR), each had a strong effect on how engaged respondents were in the research process, how free they felt to discuss topics openly and meaningfully and how creative they were in ideation and thought. We also found that anonymity has real value in research for self-expression, ideation, creativity and understanding the projected self, but less so for authentic real-world behavior of the real self. It helps us understand what could be, rather than what is. Looking at the effect of the VR mediums specifically, they have huge potential to attract new audiences to research, to engage them in a fresh and stimulating way through avatar self-expression, contextual immersion and co-creation with an infinite array of digital assets inside an unlimited range of settings.

There is tension between anonymity and authentic self-expression: anonymity is not “on or off”, but a gradient. Anonymity can empower vulnerable people by making them feel safe and provides a powerful new way of discussing sensitive topics and finding new perspectives. However, anonymity can also encourage a respondent to explore the fringes within themselves. Empathetic moderation is strongly affected by the platform

and level of immersiveness: anonymity means the moderator must work harder, starting from scratch with no preconceptions. Immersive VR is more natural to moderate than online video, but in-person has the strongest ability to create empathy and trust between moderator and respondents.

In connection with the other two experiments in this paper, looking at the same three key areas of qualitative research of exploration, evaluation and ideation, we can conclude that VR plus anonymity is best for exploration and ideation—the setting stimulating respondents’ sense of creativity and wonder and allowing them to express themselves more deeply and imaginatively. This is less so for evaluation, especially of concrete stimulus where the medium had less benefit. The real opportunity for the future, which will happen with virtually embedded GenAI tools and research friendly templates, will be combining these three areas of exploration, evaluation and ideation into a consumer-led co-creation environment with real-time development of concepts for innovation and communication.

SYNTHETIC RESPONDENTS PILOT: IS THERE DEPTH AND EMPATHY IN THE SYNTHETIC?

The coupling of traditional methodologies with innovative technological advancements, such as AI, can potentially yield a comprehensive understanding of consumer behaviors and preferences. In this context, we undertook an initiative to stress test the use of synthetic respondents—AI-generated profiles or “twins”—in gauging market reactions and ideation in the field of women’s health, specifically focusing on the menstrual cycle. This process was structured into a three-stage pilot, each stage meticulously designed to serve a distinct purpose, while collectively ensuring the integrity and utility of the synthetic respondent approach. As with our AI moderation research, we wanted to evaluate



Coplan and Goldie's conceptualization of empathy. In this research, we were mostly interested in a synthetic respondent's ability to apply affective matching, i.e., demonstrating genuine emotional responses. We wanted to test its ability to reflect emotions in a natural way, gleaning both the functional and the emotional.

Creating a unique digital twin for each human counterpart

The first stage of our 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. The project delved into the multifaceted experiences of women at different phases of the menstrual cycle: pre-menstrual, menstrual and post-menstrual. It aimed 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 understanding of the respondents' responses and preferences, which is critical for creating authentic synthetic profiles.

In the second stage, we leveraged the Ipsos unique approach to engagement in online communities to rapidly re-engage the same cohort of women. The objective of the follow-up activity was to delve deeper into their feelings, needs and challenges associated with each menstrual phase, and to solicit their active participation in ideating potential food and beverage solutions. By guiding the respondents through the process of conceptualizing products tailored to their stated needs and subsequently evaluating these ideas,

we gained actionable insights into the practical applications of their input. Concurrently, and as a third stage, we introduced synthetic respondents, individually generated through GenAI models informed by each human respondent's data collected in stage one. In essence, the AI twins were designed to simulate how the human respondents from stage one (mirroring their demographics, behaviors and preferences associated to femcare) would likely answer. By subjecting these synthetic twins to the identical set of questions and tasks as their human counterparts, spanning exploration, ideation and evaluation, we aimed to assess the degree to which AI can replicate human responses in the context of market research. Additionally, the iterative nature of this pilot, with its focus on exploration, ideation and evaluation, was instrumental in assessing the creative and analytical capabilities of the synthetic respondents. By comparing the originality and feasibility of the product concepts generated by both humans and AI, we can discern the potential of AI to not only understand existing consumer needs, but also to contribute to the innovation process in a meaningful way.

The following sections delve into the findings of the three-stages:

1. Exploration, where the physical and emotional needs are probed and compared between the human and AI respondents;
2. Ideation, comparing the generation of new product ideas of each;
3. Evaluation, where new product concepts are compared by each.

In the evaluation, different goals may exist. For exploration and evaluation, the accuracy of the AI is key, which ideation could either seek replicating what human respondents would produce, or the



ability to generate new and useful ideas, whether similar or different from the human respondents.

KEY FINDINGS

A reflection of the central tendency and the rational

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, but it will be seen in the findings that the richness of responses is also related to the accuracy here. Synthetic respondents correctly identify key themes, though generate fewer on average: real and synthetic members overlap on some of the key themes in each of the three stages of the menstrual cycle. That said, humans on average provide 72% more themes during exploration (see Figure 3).

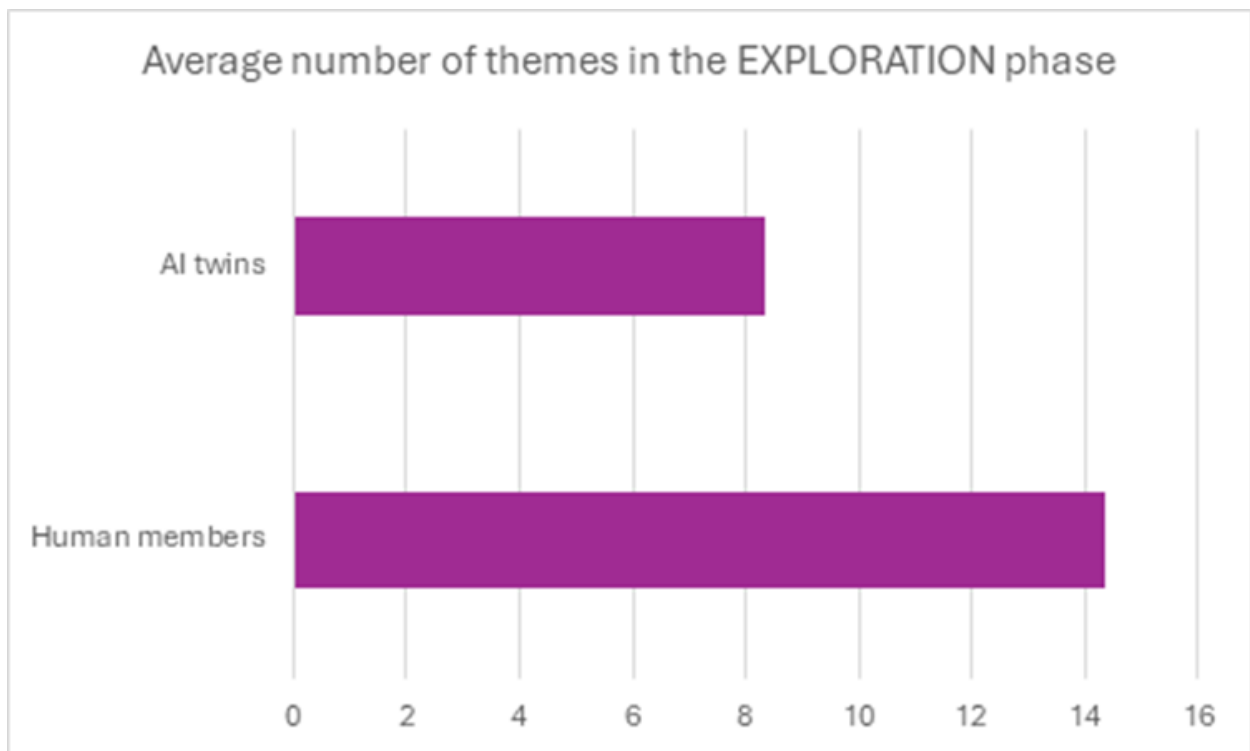


Figure 3



- Synthetic respondents synthesize insights: There is less variation in individual responses. While real community members individually describe the whole range of symptoms and experiences, their synthetic twins put emphasis on just one or two.
 - Synthetic respondents are more conceptual: Whereas real members provide concrete examples of how their pre-menstrual cycle is affecting them, their AI twins take a high-level approach.
 - Synthetic respondents are more thorough in identifying all the different physical symptoms during menstruation: While real members put emphasis on one or two symptoms they recall more readily (abdominal and back pain), synthetic respondents are unable to tell which symptoms are more relevant and will list other minor symptoms, such as dizziness or anemia, as equally important.
 - Synthetic respondents don't seem to grasp the counter-intuitive: Whereas real members report a sense of emotional relief when their period arrives, their AI twins keep insisting on the negative (irritability, depression, etc.). Synthetic respondents don't seem to grasp the counter-intuitive: Whereas real members report a sense of emotional relief when their period arrives, their AI twins keep insisting on the negative (irritability, depression, etc.).
 - Synthetic respondents don't seem to grasp the relevance and effect of social factors on behavior: Whereas both AI and real members speak to their preference for less physical effort during the period, real women put a stronger emphasis on avoiding important errands and social engagements. AI had very functional solutions to relieve certain symptoms (things only experts would highlight), but missed most of the social implications, i.e., hiding your tampons and/or sanitary pads at the bottom of your bag, trying to control mood swings while with friends, fearful of leakage in public, etc.
 - Synthetic respondents tend to be less specific when answering novel questions: Their answers are generic if their training data doesn't specifically address topics in the new set of questions. For example, synthetic respondents provided much less detail concerning the use of sanitary products during menstruation, whereas real respondents place a strong emphasis on managing menstrual blood and preventing leaks, indicating a high concern for menstrual hygiene.
 - Synthetic respondents are less able to transmit the full range of emotions at play: Although they are apt at identifying the correct emotional dimension, they are less specific when it comes to relevant implications. For example, while both real and synthetic respondents describe the post-menstrual phase as a naturally comfortable period that requires no special coping strategies, real members add more nuance. For instance, they express a feeling of refreshment and motivation, with an emphasis on increased activity and exercise, as well as a focus on diet and health during that stage.
- The exploration phase of our pilot study revealed that while synthetic respondents are capable of pin-pointing key themes in qualitative data, they fall short in several critical areas when compared



to real respondents. Synthetic responses lack the detailed anecdotes, prioritization of experiences and nuanced understanding of emotional and social factors that human respondents naturally provide. This deficit is particularly evident in their inability of affective matching through conveying the full range of emotions and the specificity required when addressing novel topics. Real respondents offer a textured, vibrant tapestry of insights, while synthetic respondents present a more uniform and less nuanced fabric. These findings underscore the limitations of synthetic data in capturing the rich complexity of human experiences, which is crucial for effective market, category or segment exploration.

Initial ideation might be better with synthetic, but struggle with emotional context

Moving from the exploration phase, the second aspect focuses on creativity and respondents', both 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. While gaps exist using the first criteria, AI does better in the second. Synthetic respondents provide more developed ideas, listing a greater variety of formats. They take a more holistic approach, as synthetic respondents list natural ingredients that contain several beneficial effects on the hormonal cycle (e.g., chasteberry, cinnamon), whereas real respondents focus on one or two foundational and generic ingredients (e.g., iron, vitamins). All product ideas are given a name, often chosen with high level of creativity.

However, unlike real members, synthetic respondents focus less on the emotional benefits of the products and more so on the functional benefits. Synthetic ideas tend to have a strong reason to believe (RTB) and functional benefits. Real respondent ideas tend to have a strong insight, and show a much clearer focus on how the product affects the mood. Synthetic respondents here also show a greater tendency towards synthesis, whereas real members hold a distinctly personal perspective when generating new product ideas. Real respondents provide a more empathetic perspective, focusing on products that directly address menstrual discomfort, and are tailored to individual needs, suggesting a more personalized approach. AI twins generalize the experience, offering a more one-size-fits-all solution that might resonate with a broader audience.

The ideation phase of our study highlighted a distinct difference in the creative outputs of synthetic and real respondents. Synthetic respondents excelled in developing well-rounded product ideas with strong functional appeal and a systematic approach to addressing menstrual health. However, they fell short in capturing the emotional depth and personalized nuances that characterize human ideation. Real respondents infused their ideas with personal insights and emotional benefits, crafting products with a clear focus on mood enhancement and individualized care. These findings suggest that while synthetic respondents can contribute valuable breadth to the ideation process, the depth and personal relevance of real human insights remain crucial for truly resonant market, category or segment innovation.



Rational evaluation will likely not be enough

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 focus on new ideas that 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.

Synthetic respondents seem to be more analytical: their preferences are informed by the ingredients and their known benefits (“Furthermore, the use of adaptogens like reishi mushroom and ginseng shows a thoughtful consideration of health in the product’s formulation”), whereas real members also focus on taste and texture (“It’s the most rewarding and delicious chocolate bar”). In fact, it

seems to be the key variable that influences their preference among the distinct products. In fact, it seems to be the key variable that influences their preference among the distinct products. In other words, the choices of the synthetic members are purely mechanical: if A causes effect B, and effect B is important to me, then I must prefer A above all other options. Synthetic respondents are also more positive in their assessments: real members don’t shy away from expressing their hesitations, doubts or questions regarding their chosen idea (“The true nature of Blisberry is a bit confusing, but I think I can get used to it with the chocolate I usually eat, and I feel grateful that my usual snack time leads to femcare”); whereas synthetic respondents are overwhelmingly positive regarding their chosen idea (“It’s not just a sweet treat, it’s a companion that offers support during a time when many women need it”).

Once again, humans generate more themes: real members provided approximately 33% more themes than their AI twins (see Figure 4).

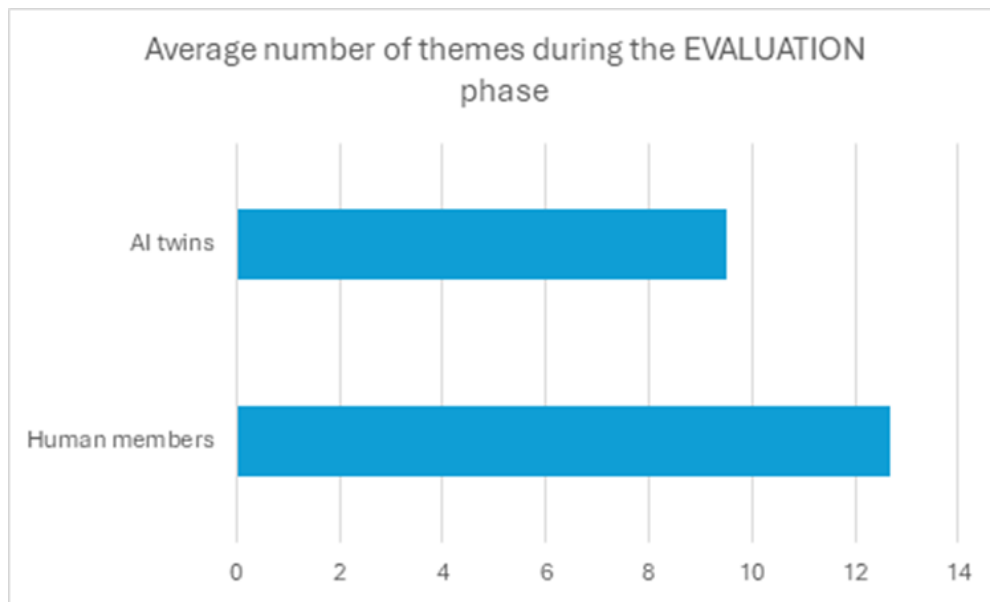


Figure 4



Preference is most noteworthy, rather than the general views. When given a clear task to choose an option that is not explicitly grounded in the

existing training data, the results of each group significantly deviated from one another.

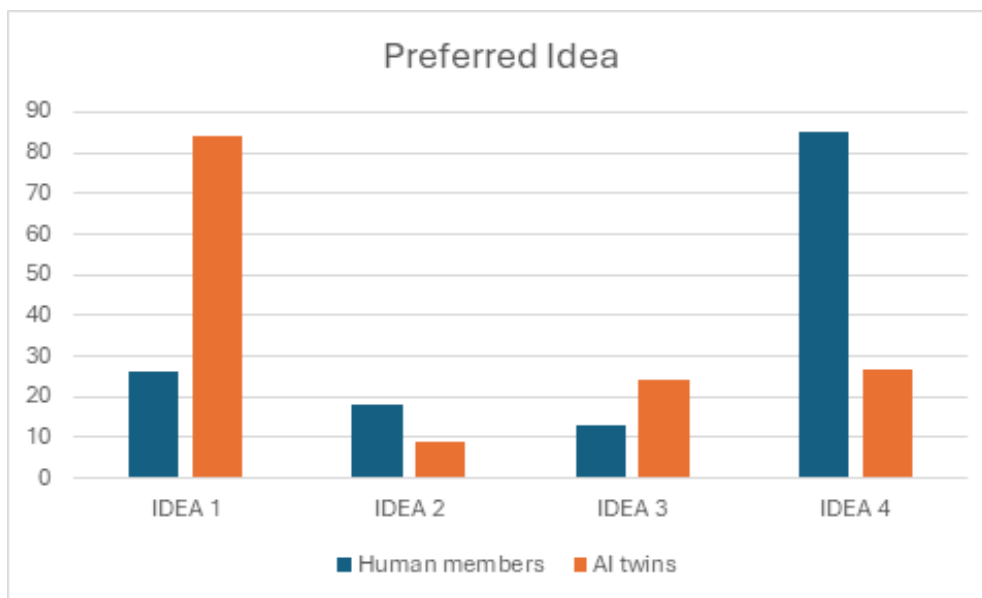


Figure 5

The evaluation stage of our study demonstrated that synthetic respondents are adept at an analytical assessment of product benefits, but they lack the nuanced judgement and sensory consideration that real respondents bring to the table. Real members' feedback was rich with personal preferences, tactile experiences and candid hesitations, offering a multidimensional perspective on product appeal. This contrast underscores the synthetic data's limitation in capturing the full spectrum of human experience, which is vital for a well-rounded evaluation in market, category or segment analysis. The depth and authenticity of human feedback are indispensable for truly understanding consumer preferences and crafting products that resonate on both a functional and emotional level.

Synthetic respondent-pilot: Conclusions

The findings indicate that AI can replicate human responses to a degree, identifying key themes and providing synthesized insights that are in line with human responses. It can provide only a sketch of human preferences and behaviors, lacking the color and texture of a vibrant painting: it consistently obviated emotional nuances, the prioritization of symptoms and experiences and the grasp of counter-intuitive or complex social factors. That said, the femcare category is an emotional one.

Where AI really shines is initial ideation: the ability of synthetic respondents to generate a variety of developed ideas with strong benefits and RTBs, suggests that when paired with a human



moderator, AI can be a valuable tool for ideation during early-stage product development. However, AI still struggled to get the insights exactly right, with its stronger focus on the functional versus emotional dimensions: real respondents provided richer detail, demonstrated a stronger emotional connection to the topics and offered more personalized and contextually relevant insights. The pilot's topic, femcare, is an emotional one. In less emotional categories, more rational synthetic responses might be warranted. However, even with these categories, this should only be considered, for now, in initial exploration and ideation. In other words, generative AI can serve as a tool for generating a broad array of ideas for consideration and testing, setting the stage for human creativity to refine and develop these concepts further. It is important to state that we would expect some overlap between human and AI-generated ideas, but AI shouldn't be expected to generate the same ideas as humans. Going to the evaluation of specific new concepts, explicitly beyond the scope of the training data, we found further deviations of the results.

While AI offers promising avenues for enhancing market research, it should complement, rather than replace, human insights. The consistent depth of themes, ideas and concepts that human respondents generate when compared to synthetic respondents should give researchers and their clients pause. What qualitative research should deliver are themes, ideas and concepts that surprise us. They should be a diverse set of emotional stories coming from individuals. These stories should trigger empathy and a desire to understand why. This pilot showcases how there are consistently deeper results from human respondents (for now). As we move forward, the optimal approach in market research will balance the efficiency and scalability of AI with the nuanced and empathetic understanding provided by human analysis. This synergistic combination promises to illuminate the most accurate and insightful path for market research and where each

type of data, primary collected directly from humans and AI generated synthetic, will add the most value in the data ecosystem. We do see a future with a hybrid data ecosystem, and while this is one use case of synthetic data, it does highlight some of its current promise and limitations, as well as how we can evaluate it to be most valuable for providing actionable insights.

OVERALL CONCLUSIONS OF THE THREE RESEARCH PILOTS

Our aim with these three studies has been to temper our instinctive enthusiasm for AI and VR with an expedient dose of pragmatism and understanding. By bringing together experiments from three pillars of qualitative research: the moderator, the setting and the respondent, we can triangulate findings to gain a fresh perspective on what value these new technologies can bring us in the industry.

AI opens up new opportunities and new use cases

While AI moderators excel in reaching a broad audience quickly, they lack the depth of human expertise. They are ideal for simple probing, short surveys, digital diaries and online community engagement. Currently AI moderation falls short in providing the nuanced understanding and empathy that human expertise offers. Synthetic respondents can be leveraged for initial ideation (followed up by human evaluation). Synthetic respondents can also be explored as a small part of sample blends to enrich human data with the expectation that whatever is generated by the synthetic will skew to the central tendency. Virtual environments and anonymity are good for ideation and exploration, because they stimulate a sense of creativity and wonder. The mask of anonymity creates a platform for richer self-expression. For some that equates to a freedom to talk about sensitive or personal topics, but for others, it's a



show where they can be someone new. Both of those are authentic in different ways. Currently avatars are a barrier to affective matching between moderator and respondent, which means they can miss cues and misinterpret responses.

It's easy to be deceived

So, what do we learn when we look at all these factors together and consider what happens when we replace human moderators with AI, when we rely on synthetic respondents for insight and when we play with context and anonymity? Firstly, it is astonishing what AI can already do—how real an AI bot sounds and how convincing synthetic insight reads. However, maybe we are deceived by our astonishment; we are duped by how real something sounds that we believe it's actually real. The mimicry is so good—like the stochastic parrot—we believe the surface is a reflection of what lies beneath. The bot, especially, sounds so lifelike—it “strokes” you with its humanlike language and intonation, it mimics reflection, conversation and perspective—but only for so long. It cannot keep this up for long without the lack of real social contact beginning to dilute the chemistry between moderator and respondent. “You totally get me!” is soon replaced by a feeling of disjointed dialogue.

Conversational quant is not a qualitative conversation

An experienced moderator sees the discussion guide as exactly that—a guide. They can meet the objectives of the project without needing to ask specific questions in a particular order. Qualitative insight comes not from the respondent answering the questions, but from the moderator questioning the answers. This ability to improvise, to pick up unexpected threads, is where the insight often lies. Going off-piste from the guide means discovery of new territory. The skilled moderator is able to shape the discussion more like a jazz

composition than a predefined performance. The bot tends to follow the script, sticking to the predefined questions. Although it sounds like they are listening in an empathetic way: “That’s awesome! Your passion for Star Wars really shines through!”, in reality that does not feed into the discussion or change the direction of dialogue in any way. It is just a response to the answer to the question, it does not question the answer to create new questions. It is essentially conversational quant. A conversational questionnaire has its uses, but should not be confused with qual and what qual can achieve.

In many ways, the AI bot acts like a junior moderator doing it for the first time: they miss opportunities to probe deeper, to pick up unexpected tangents, to trust their intuition. They spend too much time looking down at the guide, and not enough time reading the room. They have not yet developed a “nose for an insight”—that intangible yet crucial qualitative skill that separates the informational from the insightful. We do expect AI moderation to improve, but just like the first time you let a child play with the neighbor across the street, a parent will walk them over. In the future, we might let the child cross the street on their own without watching, but the time for this with AI isn't today.

Don't sacrifice the fringes

Similarly, with the digital twins, the AI reads convincingly; it gives us information that mirrors the consensus within the community. Nonetheless, it lacks an element of empathy by being too conceptual, continually synthesizing and lacking the nuance that the human members add. It represents centrist findings—again like a junior researcher—and pulls insight into the middle ground. This has a role in research and is not without value (especially as it's “free”), but is not the role of qualitative. Qual discovers the unexpected, unearths the unseen, the unsaid, the unheard. A



practiced human moderator's ability to ask non-obvious questions often leads to more meaningful insights. We look for the fringes, the edges and individual experiences beyond the masses, which is fuel to inspire our clients to be distinctive in an increasingly cluttered market. That is the power of one—one person's story that changes how you see things. If you sacrifice the fringes, you're just left with the middle ground. There is a vast spectrum of color in the world, so why settle for grey? AI holds enormous potential for our industry, but it runs the danger of vanillarising insight by being applied in the wrong places at the wrong time. Virtual and anonymous settings also provide the ability to leverage the fringes by attracting new audiences to research. VR plus anonymity is best for exploration and ideation—the setting stimulating respondents' sense of creativity and wonder and allowing them to express themselves more deeply and imaginatively. However, this is less so in the case of evaluation. It helps us to understand what could be, rather than what is.

The ever-evolving AI and HI dynamic

AI does have many exciting applications, especially in areas such as open-ended surveys, longitudinal diaries, communities and curation. For ideation, it's your endlessly enthusiastic verbose best friend. Rather than trawl through endless PowerPoints, get your AI to give you an overview. It will help you know things, but not necessarily understand them and certainly not feel them. We should be cautious of simulating insight. AI should complement, rather than replace, human insights. The depth, empathy and contextual richness that human respondents bring to the table are irreplaceable assets in understanding consumer behaviors and preferences. The dynamic between human intelligence and artificial intelligence will evolve rapidly in the research world from now on in; AI might do things we haven't even thought of yet, but for now we need to ensure balance between the

efficiency and scalability of AI with the nuanced and empathetic understanding provided by human analysis.

Have we tempered our enthusiasm? Quite the opposite—there's so much more to explore and understand with AI, and we'll never stop experimenting. We can celebrate its potential for instant ideation, timely probing and always-on engagement. Nevertheless, we now know more about its current limitations; how it moderates like a novice and misses the non-obvious; the way it sacrifices the vibrant fringes and pulls insight into the middle ground, where it's harder to find distinct opportunities for our clients; how it lacks the nuance and "nose for an insight" that distinguishes the expert human qualitative researcher.

For now.

ABOUT THE AUTHORS



Ajay Bangia

Ipsos UU Malaysia

ajay.bangia@ipsos.com



Rollo McIntyre

Ipsos UU UK

rollo.mcintyre@ipsos.com



Jim Legg

Ipsos US

jim.legg@ipsos.com

