IPSOS VIEWS

THE NEXT ERA OF B2B COMMUNITIES

How community-trained digital twins can make your data go further

Maya Ilic
Oscar Bravo



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. Our clients benefit from insights that are safer, faster and grounded in the human context.

#IpsosHiAi

The reality of B2B research

B2B research is entering a turning point. In today's rapidly changing business landscape, companies seeking to understand business decision-makers and other professionals need insights faster than ever before. But recruiting high-quality, credible participants for traditional research takes patience and investment, as senior decision-makers value their time.

To increase speed and improve access to professional audiences, many brands are turning to research communities that bring professionals together in ongoing conversations. Here, decision-makers are recruited and vetted once, and their engagement is deepened over time through multimodal research activities.

The community gives brands the chance to observe long-term changes in their audience, as well as conducting quick research. B2B communities also benefit participants by providing them with the opportunity to engage with each other.

In today's rapidly changing business landscape, companies seeking to understand business decision-makers and other professionals need insights faster than ever before.



Why online communities unlock B2B research potential

Thriving online communities of business professionals already exist outside of the research industry — forums, professional networks, and customer communities where decision-makers actively share challenges, solutions, and real-world experience.

Because communities tap into this intrinsic motivation to connect with others[1], people in market research communities are more willing to participate, as engagement creates value for them as well. These research environments give them a space to exchange ideas with peers, reflect on shared challenges, or contribute to discussions that feel relevant to their professional world.

INSIGHT **COMMUNITIES**

Synthetic approaches using research communities may make it economically and operationally viable to pursue questions that were previously too expensive, too slow, or too intrusive to ask with real-world samples alone.

ONE EXAMPLE: Ipsos built a long-term, always-on online community of 750+ hard-to-reach decision-makers across four markets to help a Fortune 10 tech client inform product development and communications. Through ongoing discussions, live chats, quick-turn surveys, and co-creation, the online community delivered 48 projects in the first year and continues to provide rapid, high-quality feedback and deep insight into motivations, pain points, and use cases, directly feeding into executive-level decision-making.

The lesson is clear: B2B research is highly effective when embedded in a community context. Even so, insight communities face limits. They still depend on human availability, motivation, and time. The real opportunity emerges when you ask: What if you could extract the richness, authenticity, and diversity of community-generated data without relying on repeated human participation?

The Next Era of B2B Communities



How B2B digital twins can make your data go further

Synthetic approaches using research communities may make it economically and operationally viable to pursue questions that were previously too expensive, too slow, or too intrusive to ask with real-world samples alone.

Digital twins — AI-powered synthetic representations of individual people, trained on rich personal data and deployed to simulate human behaviour — arrived with enormous promise. The logic was seductive: if you could build an accurate digital replica of a person, you could ask it unlimited questions without burning out the original. You could run experiments, test ideas, and iterate endlessly with no fatigue, no recruitment bottlenecks, no time constraints. For B2B research, the appeal was even clearer: finally, a way to understand complex business decision-makers at scale.

At Ipsos, this promise has been taken seriously from the start. Our foundational paper on this topic, *Synthetic Data: From Hype to Reality*[2], established the conceptual and methodological ground on which subsequent twin development has been built, mapping the conditions under which synthetic data can be trusted, where it adds genuine value, and where it falls short. It made it clear that synthetic data is not a single entity; its usefulness depends entirely on how it is generated, what it is trained on, and what it is asked to do.

Which is exactly the question that this pilot set out to address. As online community practitioners began testing digital twins in real research settings, a more fundamental challenge emerged. Given the range of data types available in online communities, what kind of data should be fed into a digital twin to make it useful, and for what purpose?



The setup: three data types, one audience

To understand which data makes digital twins functional for B2B research, we needed a real test case with a specific, niche business audience — people with genuine expertise, real challenges, and meaningful business stakes. In Mexico, independent stores such as neighbourhood shops, small grocers, and kiosks control over 70% of mass-consumption product transactions across the country's informal and formal retail landscape, making them indispensable to any brand's growth strategy in emerging markets.

Ipsos operates an always-available online community of 1,000+ *"tenderos"* (neighbourhood store owners) across Mexico, providing continuous engagement with this essential retail channel. This online community-based approach fundamentally changes the research equation. Instead of traditional recruitment methods, brands can connect with store owners faster, at lower cost, and through integrated research methodologies, such as surveys, online community conversations, and depth interviews, all within a single research ecosystem.

For our digital twin test, this online community became the ideal lab. Working with our client *Opella*, we collected training data from tenderos[3] that was directly relevant to a real business category: over-the-counter (OTC) pharmaceuticals. Three different types of training data were used to generate the digital twins:

01

Survey data: demographic and profiling information, plus responses to structured questions about their business, challenges, preferences, and strategies within the OTC category, without any open-ended responses.

02

Community data: posts and conversations from the online community where the tenderos discussed their real-world challenges related to the OTC category, layered on top of the survey and profiling data, as well as additional information from different engagement activities conducted over time.

03

Interview data: in-depth, conversational interviews capturing the nuance, reasoning, and context behind their decision-making, again added to the survey and profiling layer.

We built 30 twins per condition, one per tendero, 90 in total. We then asked all three sets of twins the same qualitative questions designed to gauge whether they transmit emotionality and nuance, if they can surface contradictions and tensions the way humans do, and how grounded they are in real business context.

What our pilot revealed: which data builds the best B2B twins

How we measured: content analysis and category deep dives

We used a two-part comparison methodology, with survey-trained twins as the baseline condition — the starting point from which the contribution of adding qualitative data could be assessed.

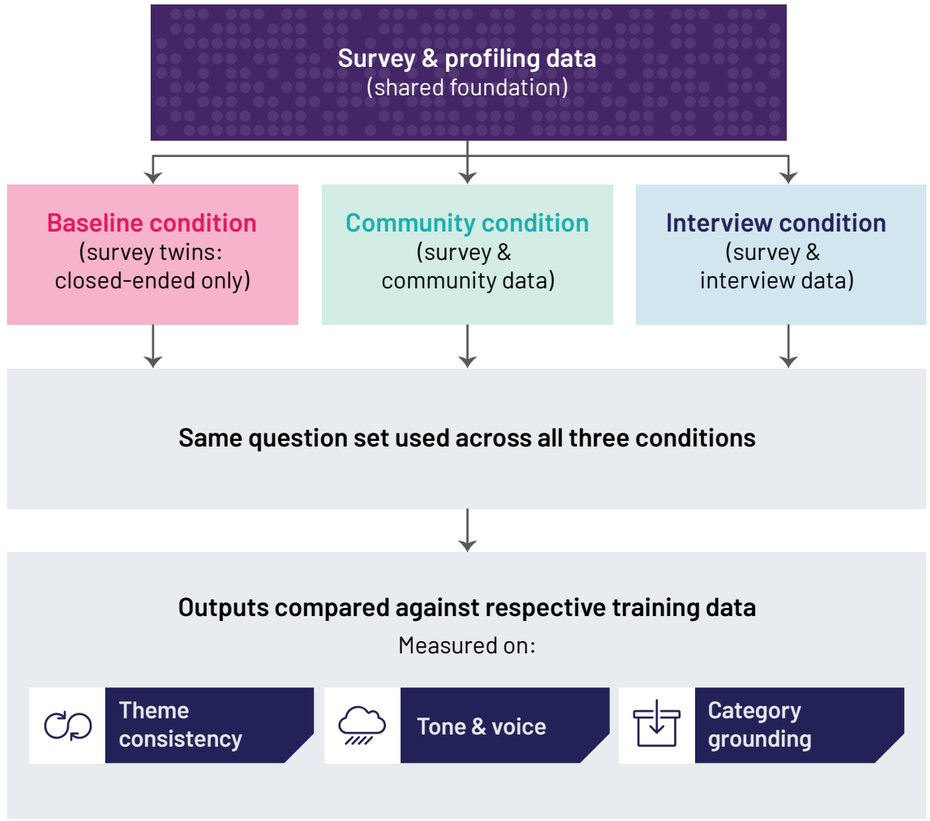
First, we ran content analysis to measure theme consistency — the proportion of themes in each twin's output that directly matched themes present in its training data (Matched Output Themes / Total Output Themes), as coded by human analysts. A theme was counted as matching when it referred to the same category topic (for example, inventory management, seasonal demand, or competitor pressure) in both the training data and the twin's output. Theme consistency captures how faithfully a twin reproduces what it was trained on.

We then assessed consistency of tone and sentiment between the training data and each twin's responses. We considered whether the emotional register, the degree of frustration or confidence, and the practical versus aspirational framing of the output matched what tenderos had expressed. In qualitative research, tone and sentiment are not incidental — they carry the meaning that makes insight actionable for communication strategy and brand positioning.



> **In qualitative research, tone and sentiment are not incidental — they carry the meaning that makes insight actionable for communication strategy and brand positioning.**

Figure 1: Calibrating the inputs: which data builds a useful B2B twin



Source: Ipsos

A note on methodology: this study measures consistency between twin outputs and their training data, rather than comparing twin outputs directly to human responses. The question of how closely twins mirror their human counterparts has been explored in previous Ipsos research, *Conversations with AI Part V*[4], and that earlier work provides the foundation on which this pilot builds. Here, the question is different and more applied: given that foundation, which type of training data produces twins that are most faithful to what they were taught, and most useful for qualitative B2B insight?

To answer that question, survey-trained twins serve as the baseline condition that represents the most common and accessible starting point for digital twin work. Community data and interview data are then tested as additive layers, to understand what each contributes over and above that baseline. All questions posed to the twins were kept closely related to topics covered in their training data. Throughout, the test is a qualitative one: we are asking which data produces twins that sound and reason like the people they were built from, not which data best predicts quantitative outcomes.

Baseline: survey twins - broad coverage, shallow insight

Survey twins showed 62% theme consistency with their training data. The survey layer here was closed-ended only — structured scales, categorical selections and fixed-format questions — so it contained no human language for the twins to draw on. This is distinct from survey approaches that use AI to elicit or process open-ended responses, which would introduce a different quality of data.

KEY INSIGHT: closed-ended survey data gives you a thematic starting point, but with no examples of human language in the input, the LLM fills the gaps with generic elaboration. More output themes is not a sign of richer twins — it is a warning sign.

Layer 2: community twins - grounded output, authentic voice

Community-trained twins demonstrated 72% theme consistency — 10 percentage points above the survey baseline, and qualitatively different in character. Their training data contained 22 themes; outputs produced 16. Instead of padding their responses with outside knowledge, the twins focused on the questions asked, synthesising the most relevant qualitative data rather than drifting into generic territory.

In response to category-specific questions in the OTC sector, they identified tenderos' key themes and provided detailed information on specific categories: differentiating between pain relievers, antihistamines, digestives, and throat lozenges; distinguishing reliable generics from premium brand positioning; and articulating the dynamics of temporada de gripe (flu season) and allergy seasons as distinct inventory moments. While actual tenderos in the training data mentioned "pain relievers" and "seasonal items" as broad categories, the twins elaborated within the same category logic.

Whether every elaboration faithfully represents what tenderos would say if pushed further is a question for future validation work. However, there is a clear difference between the two outputs in terms of register, voice, and actionable specificity. While a survey twin produced "OTC inventory should be optimised based on seasonal demand patterns and margin considerations", a community twin produced "I always make sure I have paracetamol and throat lozenges stocked before flu season — my regulars expect it." The survey twin speaks in the abstract, prescriptive register of a management recommendation — third person, category-neutral, and directionless. The community twin speaks in the first person, operational register of someone running a shop — situated, brand-aware, and grounded in the competitive dynamics of the street. This is relevant for brands developing retailer-facing communication, such as category messaging, promotional framing, and field sales narratives, which needs to mirror the language and logic tenderos actually use, not a sanitised version of it.



KEY INSIGHT: community data produced fewer but more grounded themes, constrained the LLM's generative tendencies, and preserved the practical voice of the audience — the combination that makes twins useful for qualitative B2B insight.

Layer 3: interview twins - rich data, but curation matters

Interview-trained twins showed 73% theme consistency — marginally higher than community twins, with a similar compression pattern (18 themes in training data, 16 in output). But the qualitative character of the output was strikingly different. Interview twins became markedly more aspirational and strategic, adopting a formal, solution-focused voice that did not reflect how tenderos communicate day-to-day. They sounded like consultants advising tenderos, not tenderos reasoning through their own challenges.

This is consistent with Rocha and Duarte's finding that the register in which training data is presented shapes how an LLM processes it[5]. The one-on-one interview dynamic — exploratory, reflective, aspirational - pushed the twins toward a more formal register. This is not an argument against interview data, but rather an argument about curation. Interview transcripts require deliberate preparation to be useful as twin training material: stripping out moderator framing, focusing on operational passages, preserving natural language. Online communities do this by default as the data is already text-native, peer-to-peer, and unprompted.

KEY INSIGHT: interview data is not inherently inferior. However, without deliberate curation, the register of the interview context bleeds into the twin's outputs. The structural advantage of community data is that it arrives pre-curated for this purpose.

How to read these findings

These findings are best read as guidance on how to use community data to build more useful twins for qualitative research. They show what kinds of input produce the most grounded and faithful outputs, and they do so under controlled conditions, with a specific audience, in a specific category context. The broader lesson is not that one data type "wins" in the abstract, but that data needs to be selected, curated, and combined with care if it is to generate twins that are both useful and credible.

The frame that matters: simulators, not replicas

The key issue is not the absolute percentage of theme alignment, but the relative gain from adding qualitative layers on top of the baseline. The exact number will vary with the data mix, the question type, and the method used to train the twin — and it can likely be improved with different curation choices and methodological refinements. But the deeper point is more fundamental: no digital twin is a perfect replica of a person. All twins are simulations, and the right standard is therefore not total accuracy, but useful fidelity with enough grounding, enough consistency, and enough strategic value to make the output worth using.

Qualitative dimensions

In qualitative research, tone is not decoration; it is data. A twin that identifies the right themes but strips out the emotional texture and voice of the audience will produce insight that is technically consistent but qualitatively impoverished. Tonal fidelity is a dimension of accuracy, not an additional luxury. However, when tone and accuracy diverge, accuracy takes precedence. Authentic-sounding output is not necessarily accurate, and a twin that confabulates fluently poses a greater risk than one that hedges.

The structural advantage of community data is partly about format. Interviews capture things text cannot: the pauses, hesitations and false starts through which people reason aloud. But almost none of that survives transcription, and what the twin is trained on is the transcript, not the conversation. Community data carries no such richness to lose, but what it does contain arrives intact: participants explain themselves in writing, in their own words, with context, workarounds, frustrations and local language preserved exactly as they expressed them. This is not just a content advantage; it is a curation advantage. Community data arrives in the format that LLMs process most naturally, requiring the least preparation to produce grounded twins.

Where confidence is warranted – and where it is not

Twins are grounded in what people have already said; reliability decreases as questions move beyond the territory covered in training data. Specifically:

- They are not reliable predictors of individual behaviour in novel scenarios — those questions belong with real people.
- They should not replace direct conversations for high-stakes individual decisions or relationship-sensitive issues.
- They will not perfectly reproduce idiosyncratic judgements, emotional nuances, or counter-intuitive responses of specific individuals.

Earlier work on how closely twins mirror their human counterparts[4] using community data found that they performed best where the brief was generative — surfacing options, directions, and elaborations — and least well on evaluation tasks requiring nuanced sensory or emotional judgement. The exception is where twins are built for a defined evaluative task and validated against human results: our innovation work on concept testing[6] found them to be highly predictive of concept performance. Outside of these conditions, they provide stepping stones towards future initiatives rather than verdicts on completed ones.

Authentic-sounding output is not necessarily accurate, and a twin that confabulates fluently poses a greater risk than one that hedges.



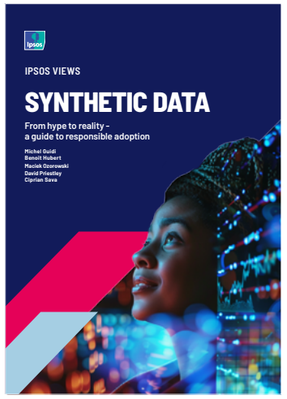
What this pilot established

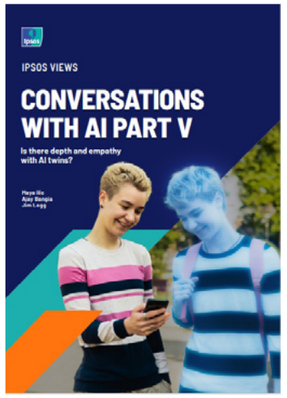
The combination and curation of data matters more than volume. Closed-ended survey data, on its own, produces generic output because it contains no examples of how people actually talk — you need human language in the input to generate human-sounding output. Community data succeeds because it combines quantitative structure with rich, unprompted qualitative context in a text-native format. Combining data types within the right methodological framework is likely the next frontier. This pilot tested them separately to understand their individual contributions, but the most powerful twins will probably draw on multiple sources, structured to minimise register interference and maximise topic grounding.

Accuracy and authentic tone must work together. Community twins maintained closer fidelity to real communication patterns than interview twins — and tonal authenticity is genuinely valuable for communication strategy and positioning work. But authentic tone without accuracy is actively misleading. The value of community data is precisely that it tends to produce both simultaneously.

Embrace the "simulator" mindset. Digital twins are advisory engines rooted in real human data that are capable of extending, structuring, and pressure-testing human knowledge, but not substituting for it. Not all questions are equally predictable. Topics grounded in observable, operational behaviour are more reliably reproduced than emotionally complex or counter-intuitive ones, as our previous research on research shows. To transition from experimental tools to trusted research infrastructure, building confidence measures into twin deployment within online community environments must become the industry standard — giving practitioners a visible reliability indicator based on underlying data density.

Further reading







Endnotes

1 Ipsos Views (2019). For the love of money? Motivation and engagement in online research communities

2 Ipsos Views (2024). Synthetic Data: From Hype to Reality

3 Note: Opella did not receive personal data from shop owners participating in the community.

4 Ipsos Views (2024). Conversations with AI Part V

5 Rocha, A. & Duarte, L. (2026). Why Your AI Persona is Lying to You: The Danger of Data Overload in Synthetic Populations. Research World, ESOMAR

6 Ipsos Views (2026). Concept Testing with Digital Twins

SEPTEMBER 2026

THE NEXT ERA OF B2B COMMUNITIES

How community-trained digital twins can make your data go further

AUTHORS

Maya Ilic
VP, Global Communities Innovation, Ipsos

Oscar Bravo
Business Director, Communities and Social Intelligence, Ipsos in Mexico

The **IPSOS VIEWS** white papers are produced by the **Ipsos Knowledge Centre**.

www.ipsos.com
@Ipsos