IPSOS VIEWS

CONCEPT TESTING WITH DIGITAL TWINS

Humanizing AI series, part four
How synthetic data accelerates innovation

Colin Ho, Ph.D
Jiongming Mu
Yuding Duan

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.

#IpsosHiAi

> 82% of leaders expect innovation to drive growth over the next three years, yet 60% believe their innovation cycles fall short.

Synthetic data is transforming market research – and this is only the beginning

At Ipsos, we have been exploring the frontiers of synthetic data boosting, imputation, fusion, persona bots, and digital twins – continually seeking out new ways to accelerate development and unlock human understanding.

In our previous publication, *The Power of Product Testing with Synthetic Data*[1], we looked at how synthetic data boosting can expedite product development cycles while providing richer insights into sub-groups. Building on this, *Concept Testing with Digital Twins* delves into an earlier stage of the innovation funnel, focusing on how digital twins can enhance idea and concept screening.

Marketing and insights professionals worldwide are feeling the pressure: **82% of leaders expect innovation to drive growth over the next three years, yet 60% believe their innovation cycles fall short**[2]. With millions of dollars riding on every product launch, brands face a difficult balance between speed and accuracy.

This paper provides readers with clear guidance on when to use synthetic data solutions to achieve innovation goals, the different types available, and how to ensure the solutions are credible and fit for purpose. We outline:

- The role of digital twins in innovation screening
- The effectiveness of different approaches
- Suitable tools for varying risk levels

For over fifty years, Ipsos has been at the forefront of consumer-centric innovation, helping thousands of brands across sectors globally launch new product and services that resonate and succeed in market.

We demonstrate how marketing and insights professionals can accelerate their innovation cycles and trust predictions – **it all comes down to how those predictions are made.**

But first, what is a digital twin?

The words "digital twins" mean different things to different industries. If you search these words, you will see many definitions – the majority of which relate to physical objects or processes in manufacturing. In marketing research, however, digital twins are virtual AI representations of real *individual* people generated from *real data* to *simulate* attitudes, decision-making, and behaviors. Digital twins allow for instantaneous answers to questions and allow for population projections *if* built from sufficiently large representative samples of humans (e.g., 1,000 digital twins to represent a country's consumers). The main benefit of digital twins: **accelerating the speed at which insights can be retrieved and hypotheses can be explored before moving on to human validation.**

Digital twins are data-driven models aimed primarily at simulation, among other uses. The process to establish the reliability of digital twins is therefore similar to those for AI predictive models that have preceded Large Language Models (LLMs). Digital twins are reliable if they accurately predict what we want to predict. Unlike previous AI models that pre-date LLMs, however, digital twins can be used qualitatively or quantitatively. Therefore, the requirements for quality of response vary by application.

This paper focuses on quantitative use cases for digital twins. For low-risk decisions where speed to insight is paramount, manufacturers may be willing to have predictions that are just "good enough". If digital twins are to be used for high-risk business decisions, then accuracy needs to be higher. In short, digital twins are not inherently good or bad. Rather, the question is whether they are good enough for what they are used for. In the next pages, we focus on the development and validation of digital twins for the specific purpose of *mass screening* of new product ideas or concepts in specific domains (e.g., food) and market (e.g., US). With the advancement of AI, marketing and insights professionals can generate large quantities of product ideas in minutes or hours. In such situations, they need an efficient way to screen through hundreds of ideas so that a smaller set can then be further developed, and more rigorously evaluated by human consumers. While we focus on this specific case, the principles laid out in this paper apply to the development and validation of digital twins, in general.

> Digital twins are virtual AI representations of real *individual* people generated from *real data* to *simulate* attitudes, decision-making, and behaviors.

Concept Testing with Digital Twins

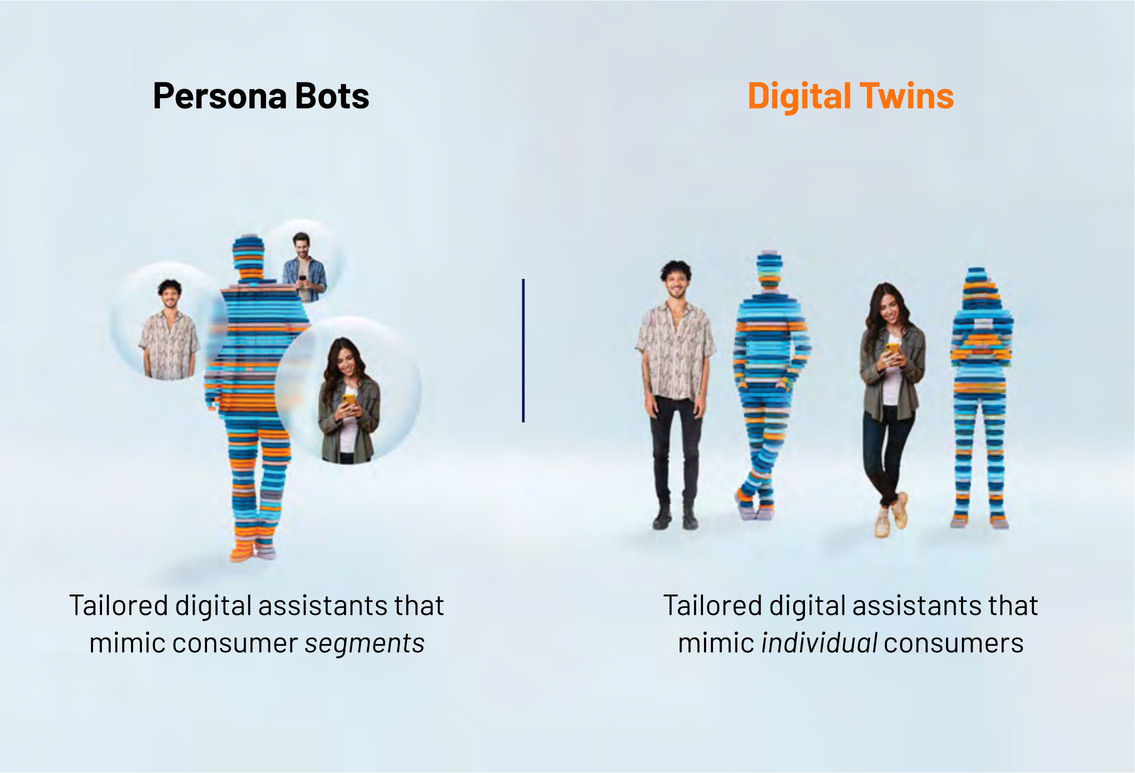
Real, Individual People

AI Representations (Digital Twins)

What separates digital twins from persona bots?

Digital twins should not be confused with persona bots. Persona bots are AI virtual representations of *groups of individuals*. To represent five consumer segments, for example, we could develop five AI personas to represent each of the segments. Persona bots are built using aggregated data (e.g., text summaries of each segment). Digital twins are built using individual-level data. The answers from persona bots are qualitative only and should not be used for quantitative projections. Digital twins are only one form of synthetic data. Data boosting, imputation, and data fusion represent other forms. These other forms, however, do not allow a researcher to ask new questions. They only serve to augment or enhance existing *numerical* data sets. We do not discuss these other forms here but direct the interested reader elsewhere[3].

Figure 1: The difference between persona bots and digital twins



Source: Ipsos

Concept Testing with Digital Twins

Know what you are trying to predict

The mantra of human intelligence and artificial intelligence (HI+AI) has been repeated so often that it risks becoming hollow and losing its effectiveness. But the interpretation of what HI+AI means is not always well understood. HI+AI is not just about having humans review AI's output. Rather, HI+AI means *weaving human intelligence into the very fabric of an AI solution.* This means incorporating learnings from previous research or experience on how humans think, feel, believe, or behave.

Two major principles from previous research are critical to follow when building digital twins. First, the **principle of specificity** comes from social psychology. This principle posits that the predictive power of attitudes on behaviors is enhanced when the attitudes are measured at a level of *specificity that closely matches the behavior of interest.*

> **HI+AI is not just about having humans review AI's output. Rather, HI+AI means weaving human intelligence into the very fabric of an AI solution.**

For example, an attitude specifically toward 'eating vegetarian salads' will more accurately predict the behavior of 'choosing a vegetarian salad for lunch' than a broad attitude toward 'healthy eating.' This principle highlights the importance of aligning the specificity of attitudinal questioning with the particular behavior or outcome being studied, thereby improving the accuracy of behavioral predictions.

Applied to digital twins, this means there needs to be a direct relationship between the information we use to build the twins and what we are trying to predict. The information used to build digital twins does not have to be limited to attitudes but should be directly relevant to what we are trying to predict. This means you need to know what you are predicting in the first place! Following this principle, we set a specific goal for our first digital twin model: predict how *American* consumers would respond to new *food* product ideas. We limited model scope to only food because attitudes and behavior vary wildly across different consumer packaged goods categories. We also limited scope to one country as food attitudes and choices differ from country to country. Fundamentally, we're training LLMs, which are the building blocks of digital twins, to answer survey questions like a representative US consumer in the scenario of buying in a particular category.

Narrowing our focus

If we adhere to the principle of specificity rigorously, we could become even more specific when predicting consumer behavior in a category. We could, for example, build a model that predicts how US consumers would respond to new *yogurt* product ideas, instead of just food product ideas in general. To build such a model, we would need to build our digital twins using American consumers' attitudes towards yogurt. But following the principle of specificity to the letter can be impractical. There is a trade-off between specificity and generalizability. A panel of digital twins built to predict acceptance of new yogurt ideas can be used to only predict new yogurt ideas, but not frozen food, snacks, desserts, cookies or any other food categories. **As such, there needs to be a balance between specificity and pragmatism.**

Earlier research by Ipsos has indicated that there are a lot of commonalities in drivers to purchase across various food sub-categories. To broaden the applications of our twin model, we decided to build digital twins using generalized attitudes towards food as a whole so that our twins could predict across different food categories. Some accuracy may be sacrificed, but as the objective is mass screening at the early stage, the breadth of application would be broader. A general food model would also allow marketers to screen new product ideas at the fringes, or disruptors of current categories (i.e., new product ideas that don't fit neatly in pre-determined food sub-categories). Having a general food model, however, does not preclude us from enhancing the digital twins with sub-category specific training data in the future.

Concept Testing with Digital Twins

Broadening our reach

The principle of specificity goes against efficiency. The holy grail would be digital twins that can be used to predict *any* consumer attitude, reaction or behaviors for any product or services. Researchers at Stanford University had such a goal in mind, albeit in the domain of policymaking and social sciences. The researchers built digital twins using data from a two-hour audio interview that spanned a large range of topics from participants' life stories (e.g., *"Tell me the story of your life—from your childhood, to education, to family and relationships, and to any major life events you may have had"*) to their views on current societal issues (e.g., *"How have you responded to the increased focus on race and/or racism and policing?"*).

With these digital twins, the researchers managed to replicate participants' responses on the General Social Survey across a wide range of social topics like civil liberties, crime, intergroup relations, health, morality, and national issues. The accuracy was 85% defined as the ability to replicate participants' own answers two weeks later. This research generated much excitement as it provided evidence that general purpose digital twins could replicate human behavior across many domains in policymaking and social science.

It would be difficult for us to achieve the same degree of success in the world of consumer insights. First of all, it is not at all practical to conduct two-hour interviews. Most consumers would not tolerate it. With compensation, some may, but the range of consumer topics that would need to be covered would far exceed those related to policymaking. Additionally, as markets change frequently, digital twins need to be updated. Long tedious interviews would make updates cost-prohibitive or practically impossible. Most importantly, the insights sought in market research need to be granular and specific. For example, will a high protein formulation be favored over a low sugar one in a new snack bar?

Training twins with general social topics does not get to this level of granularity and is not specific enough to predict how a person may respond to specific new product ideas. Development of general-purpose digital twins to evaluate ANY market behavior is not a feasible solution. All said, the model building method by Stanford Researchers to turn individual level training data into their digital twins serves as a good template and is applied in our practice. The difference, however, is that the training data we use to build our twins is focused on a very specific application.

Humanizing digital twins

To mimic real-world consumer decision-making, we leveraged knowledge from two Ipsos business lines to build our digital twins:

01 **Ipsos' Market Strategy and Understanding unit**, which uncovers what people think, feel, say, or do

02 **Ipsos' Innovation unit**, which focuses on predicting innovation success

The learnings from our Market Strategy and Understanding unit guided the information we used to build the digital twins (see Figure 2). We captured everything we learned on what drives food choices. These include the specific ingredients people look for in food packaging (e.g., low sugar, amount of protein) to food-related attitudes such as whether a person prioritizes convenience, price, enjoyment, health, or environmental concerns when making food choices. We also collected information on people's likelihood of switching food brands. People's willingness to try new food products is strongly linked with their attachment to what they are eating today. We did not collect any information on pricing as the purpose was to screen early-stage unpriced concepts. Building digital twins to accurately reflect individual price sensitivity would be a much more complicated task.

Rather than using responses to past stimuli in the database, we instead used attitudes, preferences, behaviors and habits to train digital twins – enabling the twins to generate responses like real consumers facing new stimuli that are not relying on past data. The principle is simple – yesterday's insights don't predict today's actions. This training method is critical to ensuring the twins are "future ready" instead of only regurgitating answers to old stimuli.

The information was collected using closed-end survey questions aided by AI moderation bot. This unique mix of structured and unstructured data provided a rich source of data to build the profile of each digital twin and exemplifies how human intelligence can be weaved directly into a consumer insight AI tool.

> **Rather than using responses to past stimuli in the database, we instead used attitudes, preferences, behaviors and habits to train digital twins – enabling the twins to generate responses like real consumers.**

Figure 2: Input for US digital twins for prediction of new food ideas

What category influences shape digital twin modelling?

Example: five dimensions that matter in food choice

Preferences

- Sugar, fat, protein
- Eating preferences
- Snacking routine

Among others

Attitudes

- Easy/convenience
- Value/price sensitivity
- Enjoyment/relaxation
- Health attitude
- Environmental concerns

Among others

Branding attitudes / behavior

Attitude towards current brands, switch behavior, variety seeking behavior, etc.

Purchase habits

Label reading, purchase planning, info seeking, etc.

Goals

Unmet needs and aspirations towards food

And... personal dimensions

- Household composition
- Demographic variables
- Occasions and lifestyle related to food

The knowledge from our Innovation unit guided what we asked the twins to predict. When consumers encounter a new product in the real world, they don't evaluate the new product in a vacuum. Rather, the choice is always between an incumbent product and the new offer. For example, let's say you see a new yogurt product. If you are already buying a particular brand of yogurt, then the choice is between whether you stick with your current brand or try the new product. If you are not buying yogurt today, the choice is between staying with your current food routine versus trying a new yogurt. The digital twins, therefore, were set up to predict whether a consumer would choose an incumbent product or the new product. These choices were made with key metrics that our database has shown to correlate with in-market success (e.g., relevance). Following the validated protocols on what we ask the digital twins to predict also allows the data produced be used in existing quantitative models that can be compared to our database, this makes the analytics and reporting of digital twin prediction easy to interpret by being comparable to human based concept test surveys.

This illustrates the second principle we followed when building our twins: **the principle of what makes us human.** This is not a formal academic principle. It is simply the proposition that there are traits that make us uniquely human. In addition to making choices that are relative, there are other factors that make human reasoning unique: memory constraints, distractions, fatigue, emotions, values, and context. Digital twins are not impacted by any of these. When building digital twins, we should consider these factors. The goal, however, is to make digital twins human-like, not human. We can only do so much.

Assessing accuracy and stability

To be useful, the predictions made by digital twins need to be accurate. While many think of accuracy in terms of the ability to match the responses humans would give, we need to also consider the concept of reliability: getting predictions that don't have large fluctuations is also important. Reliability means the results are relatively stable over samples or time. We need both. Digital twins can be reliable but not accurate. You get the same wrong answer every time! A twin model can also be accurate but not reliable. You might get an accurate result once in a while, but your results may vary from project to project. It would be best to have both: accuracy and reliability.

We also need to talk about person-level accuracy versus aggregate accuracy. In the world of data science, person-level accuracy is the accepted form of accuracy. But because we are building an early-stage screening tool to help brands narrow down an initial pool of new product ideas, person-level is not the most critical criteria. Person-level accuracy validation is particularly relevant when the objective is to tailor products or services to individual preferences. While important, this is not the primary goal in early-stage screening of product ideas. Aggregate-level accuracy validation becomes more relevant when the goal is to understand broader trends or make predictions at the population or group level. This approach is beneficial for businesses designing strategies or campaigns that target large segments of consumers and is indeed what is done in concept and product testing.

As such, we evaluated our US food digital twin model using aggregated accuracies (see Figure 3). We share the predictions for 12 concepts that come from a diverse set of food categories. The digital twin predictions were benchmarked to human results on Ipsos validated trial index, and the quintile in which the trial index falls. The quintile is a classification of a concept performance based on Ipsos database. A concept can be classified as High (H), Medium High (MH), Medium (M), Medium Low (ML) and Low (L). We did not require a perfect match in quintile as the twins are meant for early-stage idea screening to narrow down large pools of ideas. The ideas that are passed through this initial screening would subsequently be re-evaluated in a more rigorous manner. If a concept falls within +/-1 quintile of human evaluation, we considered that a good enough accuracy (i.e., "pass"). Among the 12 concepts, 11 passed using this criterion. Our findings provide initial evidence that digital twins can be used to predict new product success moderately well when trained with the appropriate data.

In subsequent R&D and pilot studies, we consistently found that digital twins can match human responses 85%+ when the criteria are set at matching +/-1 quintile of human evaluation. This criterion has been used since the early 2020s with another concept prediction tool used at Ipsos – InnoPredict AI. This level of accuracy is appropriate for early-stage screening that will be followed by more rigorous testing with human respondents and is widely accepted by manufacturers as "good enough" for low-risk initiatives in practical situations.

Figure 3: Aggregated accuracy examples: US digital twin model (food category)

Trial Potential Index

	Human	Digital Twins	Pass Accuracy Standard
Concept 1	M	H	No Pass
Concept 2	MH	M	Pass
Concept 3	MH	MH	Pass
Concept 4	MH	MH	Pass
Concept 5	H	H	Pass
Concept 6	MH	MH	Pass
Concept 7	MH	H	Pass
Concept 8	H	MH	Pass
Concept 9	H	MH	Pass
Concept 10	M	M	Pass
Concept 11	M	MH	Pass
Concept 12	M	M	Pass

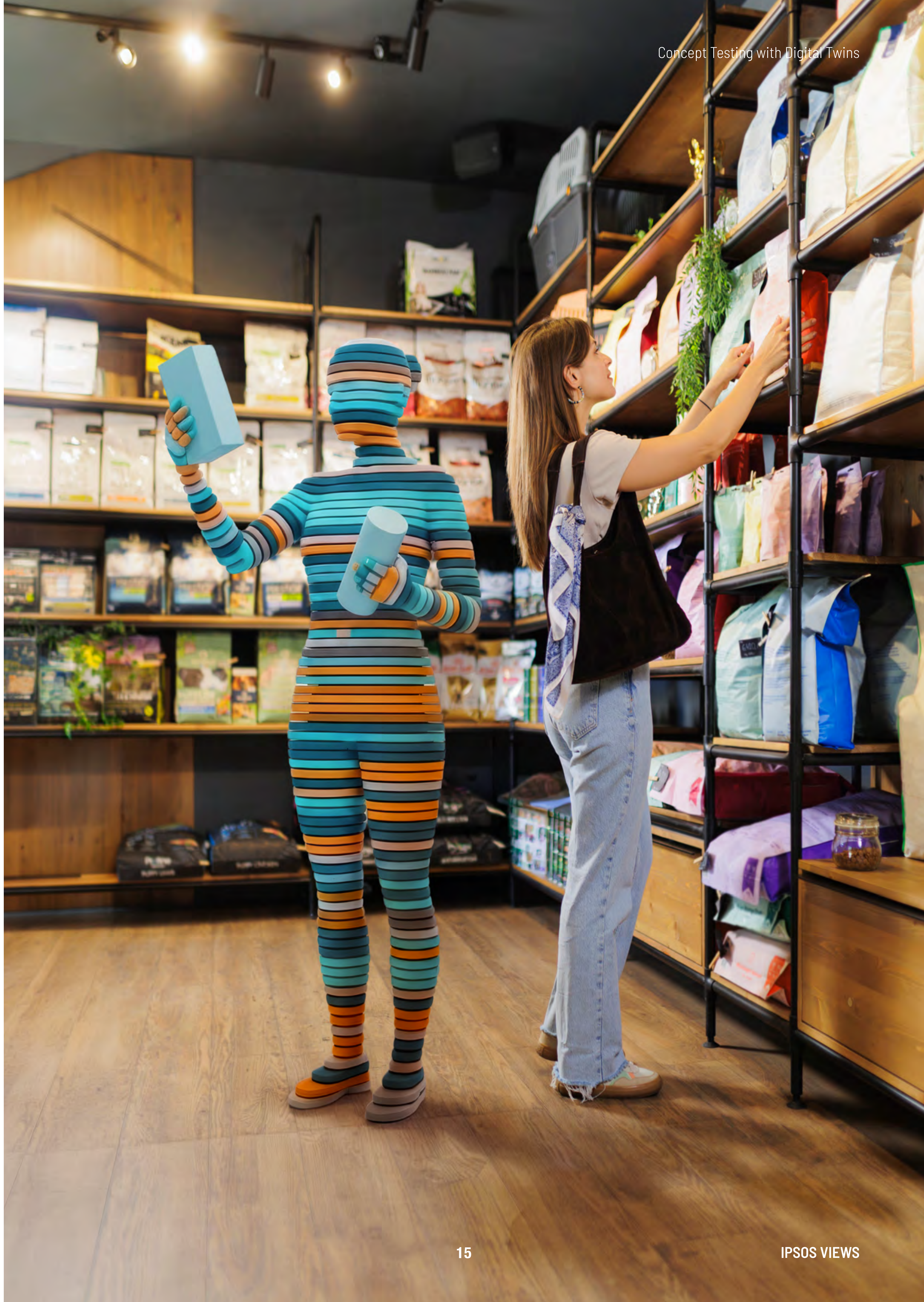
Source: Ipsos

What about variability?

To provide some perspective, we repeatedly tested the same batch of concepts among separate but equivalent human sample waves (e.g. balanced on age, gender), alongside their digital twins. What we found was that the variation between humans and their digital twins in the same wave is not larger than the variation we see between waves of human samples. Viewed in this light, digital twins offer a similar degree of stability vs. sampling error, this offers further evidence that screening early-stage stimuli with digital twins may not increase accuracy risks beyond inherent sampling risks from survey research.

One notable difference, however, is in cases where concepts are only marginally different from each other (such as minor wording difference describing an ingredient), the digital twins give answers that vary less than human respondents. This can be interpreted in a positive light that digital twins are more stable when concept differences are small. But it can also be interpreted as digital twins did not have the ability to discern small differences in concepts because the training data never got to that level of granularity while humans could. In practical business applications, this means digital twins would not be a good tool to differentiate between new product ideas that differ only slightly. Often, researchers test concepts with slight wording or content variations. Our recommendation is to use digital twins for testing the core proposition of new product ideas that are discernibly different. Digital twins should not be used to test slight differences in wording between concepts.

> **Our recommendation is to use digital twins for testing the core proposition of new product ideas that are discernibly different.**



How to choose the right agile concept prediction solution

There are different ways to quickly predict new product success. In this paper, we share a complete picture, beyond digital twins, the most popular is to build predictive AI models using concept databases to predict the success of a new product. This involves looking at new product ideas that have been tested in the past. If certain features performed well in the past, then the assumption is that these features will perform well today. Conversely, if certain features performed poorly in the past, then the assumption is that those features will also perform poorly today. These predictive models assume that knowing the past can predict the future. In categories where products don't change very much (e.g., hand dishwashing detergent), the past may indeed predict the future. For other product categories where products change substantially (e.g., think of the first mobile phones vs those we have today), the past is unlikely to predict the future. This is the primary weakness of AI models based on the past, when the purpose is to predict innovation which by definition has not existed in the past. While prediction using past data could achieve satisfactory accuracy in other realms of insights, it is the risk is high when used in the prediction of a new product's success.

Instead of using past data, Ipsos operates an innovation prediction model called InnoPredict AI that uses machine learning built on consumers' natural language responses to an innovation stimulus to predict success, by detecting themes and emotions from those answers that indicate adoption potentials. This method effectively reduces concept testing to just one question and can be scaled to any country or category. The disadvantage, however, is that it is not instant prediction as we still need to wait for human samples to respond to those stimuli.

Using persona bots to evaluate concepts can be sufficient if brands need a quick qualitative assessment or a way to optimize concepts for quantitative testing in the future. Early qualitative insights from persona bots can be extremely useful. But as persona bots represent groups of people, the responses from persona bots should not be used for quantitative projections. We summarize the different approaches in Figure 4.

Figure 4: When to use digital twins versus alternatives

*Risk	Solution	Applicability
Low	**PersonaBots** Aggregate models of groups that provide qualitative feedback on concepts	**Best for** Qualitative insights **Limitation** No quantitative projections
	Database models Aggregate predictive models that provide quantitative feedback on concepts based on the past	Stable categories (e.g., dish detergents) May fail if products innovate substantially over time
	InnoTest with synthetic respondents Simulations of individual consumers reacting to innovation concepts	Mass screening of diverse concepts Not for minor or small differences between concepts
	InnoPredict AI with humans Predictions of consumer reactions to innovation concepts with machine learning artificial intelligence	Countries/categories where digital twin panels are not yet built Fieldwork with real humans means the prediction is not instantly acquired
High	**InnoTest with humans** Predictions of consumer reactions to innovation concepts using robust, validated benchmarks	Validation test, concept optimization, testing small differences, and coherence of proposition/image/price **No limitation, the gold standard**

Risk Level for Innovation

Source: Ipsos

Future directions and caveats

This paper outlined what it takes to build trusted digital twins that simulate real consumer decisions – equipping you to evaluate competing solutions critically and leverage synthetic data for innovation with confidence. There are no short cuts. First, we need data tailored for the prediction task in mind, and that is reflective of the market and product category for which we want to make predictions. Pre-trained off-the-shelf LLMs alone cannot capture local/cultural attitudes towards specific product categories. Once developed, digital twins need to be validated. Some form of validation, even if limited, is needed to understand if digital twins can accurately predict in a specific market and product category.

The work does not end after model validations. There may be a need to update a digital twin panel if there are changes in the market that would substantially shift consumers' preferences and attitudes. Continuing with our food example, if people start believing that animal fat is good for you instead of something to be avoided, we will need to update the digital twins to reflect that. While we all desire AI to make our work more efficient, the truth is that the development of these models takes time and effort.

One area that deserves more attention — if we are to use digital twins to mass screen new product ideas — is the use of benchmarks:

- Do the benchmarks we use to denote differences between ideas for humans also hold for digital twins?
- Or do we need different benchmarks entirely?

If digital twin output is of lower variability, the differences that are considered as significantly different for humans may need to be adjusted for digital twin data (e.g., the bands that define each quintile may need to be narrower). While we have decent tools and conventions for managing variability in human respondents and to separate the signal from the noise (e.g., sampling theory, confidence intervals), these tools don't apply when dealing with AI model outputs. Finally, it is worth noting there is inherent variability and inconsistency in LLM outputs, that individual LLMs have been shown to degrade over time, and new/updated LLMs are coming to market all the time. Because of the constant change in LLMs, digital twins training will need to be refreshed and re-calibrated periodically.

Digital twins, as a method, offers a bright future of innovation research applications. While we gradually scale the method to cover more countries and categories, we are also exploring the use of twins to answer other types of innovation questions, such as concept and product feature optimization. The digital twins training can be enhanced with broader Ipsos and industry consumer data. The innovation-specific training data can also be tailored to the specific needs of our clients to make them versatile to customized research needs. We are only just scratching the surface.

Conclusion

We have sought to provide a simple yet clear overview of our approach to developing digital twins at Ipsos for idea and concept screening. While digital twins promise to transform market research, cautious deployment is essential. By integrating a foundational LLM, such as Gemini, with individual-level data, we aim to enhance the LLM's capacity to predict specific behaviors and attitudes.

To achieve this, adherence to the principle of specificity is crucial, along with incorporating the nuances of human context. It's vital to remember that digital twins are not replicas of humans; they are approximations. When behavioral decisions are driven by a limited set of rational factors, digital twins can be highly effective. However, when emotional or contextual influences are present, adding human contextualization to the model is necessary.

Moreover, digital twins should primarily be applied in low-risk scenarios, like early-stage screening, where their approximation abilities can best be utilized. There are inherent trade-offs with any technological application, and digital twins are no exception. They excel in low-risk, high-volume, screening but are not suited to final/no-go decisions. Ultimately, the future of market research isn't replacing humans — it's using AI to make human research more strategic and impactful.

> **The future of market research isn't replacing humans — it's using AI to make human research more strategic and impactful.**

Key takeaways

01

You CAN screen ideas and concepts instantly and get the right results. Once developed, digital twins are AI-powered virtual consumers that can be used to predict product acceptance in hours, not weeks. They enable instant mass screening of hundreds of product ideas — letting marketing and insights professionals focus human-centric research on the most promising concepts and accelerate time-to-market decisions.

02

Digital twin development makes all the difference. Digital twin models should balance specificity (e.g., trained using actual food attitudes and behaviors), flexibility (e.g., works across all food categories), and practicality (e.g., 15-minute survey). Additionally, it is crucial not to just simply feed any data into an LLM. Behavioral science should be embedded where needed. In our example, our critical innovation was to ensure that the twins evaluated products like humans do — comparing new options against their status quo, not in isolation.

03

Use it right, or don't use it at all. Use digital twins for what they were designed for. Once developed, care should be taken to ensure that digital twins are only used for their intended purpose. In our case, we designed our digital twins to excel at screening fundamentally different concepts. They're not designed for A/B testing minor copy changes or slight variations in wording. Ensuring proper usage prevents abuse of the tool.

To learn more about accelerating innovation cycles with synthetic data, please contact your local Ipsos team.

Methodological notes

Building and validating digital twins for concept screening

To build comprehensive foundational profiles for our US Food Digital Twins, we extracted unique viewpoints, preferences, and linguistic styles from 1,000 real US respondents using structured and unstructured survey data. We trained these models using 12 food concepts, systematically augmenting the data, altering tone, length, and structure to ensure the models were robust against varying inputs.

To accurately predict both quantitative KPIs and qualitative feedback, we fine-tuned advanced Large Language Models using state-of-the-art adaptation techniques, which ensures digital twins provide answers that resemble real humans in variability, realism of language, and positivity of sentiments.

For qualitative responses, we designed a framework featuring a preliminary "sentiment layer." This acts as a human-like "gut feeling," ensuring the twin's generated opinions naturally align with its specific persona before it articulates a full response.

When a digital twin detects that a concept has low relevance, it triggers a step-by-step optimization process. The twin reviews its personal preferences, identifies inconsistencies in the concept, and generates targeted suggestions for improvement.

Rigorous validation and optimization

To prove our approach works, we applied strict scientific validation. Using a leave-one-out cross-validation approach, we repeatedly tested the model's performance on completely unseen data. We ultimately validated the US food model against 30 distinct food concepts, achieving a 97% accuracy rate against our proprietary trial potential metric (defined as falling within +/-1 trial potential quintile).

To prove this approach is generalizable, we recently built a US pet food model, which achieved an 88% accuracy on 16 pet food concepts. Importantly, this +/-1 quintile accuracy standard is not a new AI benchmark; it is an established Ipsos standard that our clients have long trusted for early-stage human screening.

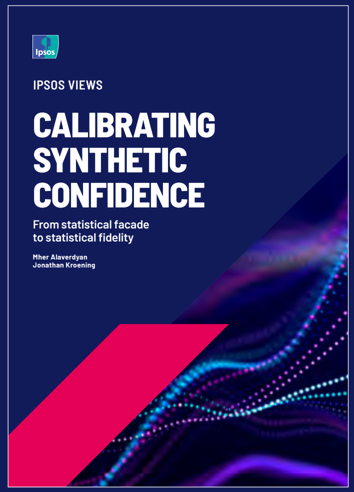
Endnotes

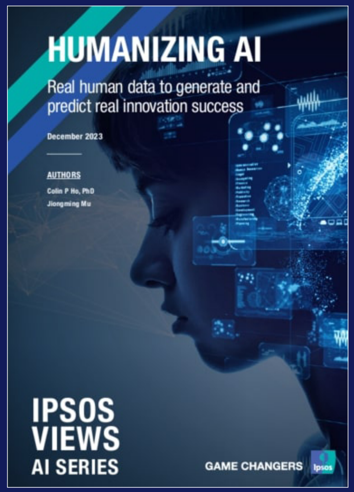
1 (2025) Ho, C. Reynolds, N. Ipsos Views - The Power of Product Testing with Synthetic Data.

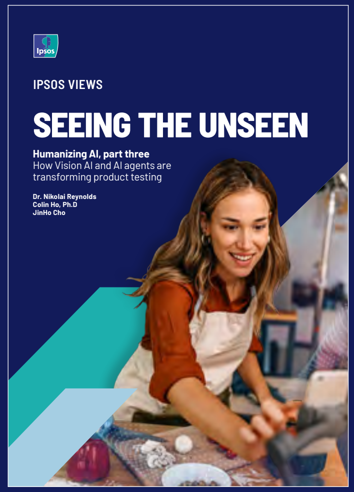
2 (2026) Ipsos and MarketLogic – Innovation, Reignited. Base: Executive-weighted sample (n=250) across CPG and adjacent sectors – 64% CEO, 24% CMO, 11% CGO, 2% Regional/Market/President/CEO - from organizations with $100M to $2B in annual revenue, fielded September 5-October 16, 2025.

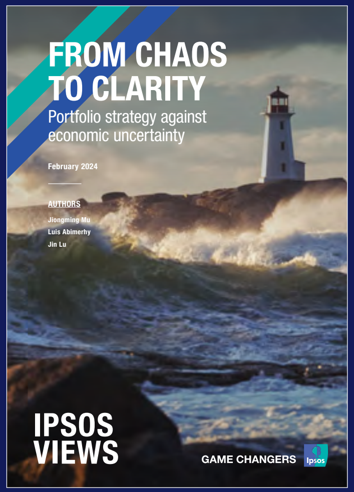
3 (2024) Guidi, M. Hubert, B. Sava, C. Timpone, R. Ipsos Views - Synthetic Data, a Guide to Responsible Adoption.

Further reading









JUNE 2026

CONCEPT TESTING WITH DIGITAL TWINS

Humanizing AI series, part four

AUTHORS

Colin Ho, Ph.D
Chief Research Officer, Innovation and Market Strategy & Understanding, Ipsos

Jiongming Mu
Global Innovation Solutions Leader, Ipsos

Yuding Duan
Data Scientist, Innovation, Ipsos

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

www.ipsos.com
@Ipsos