

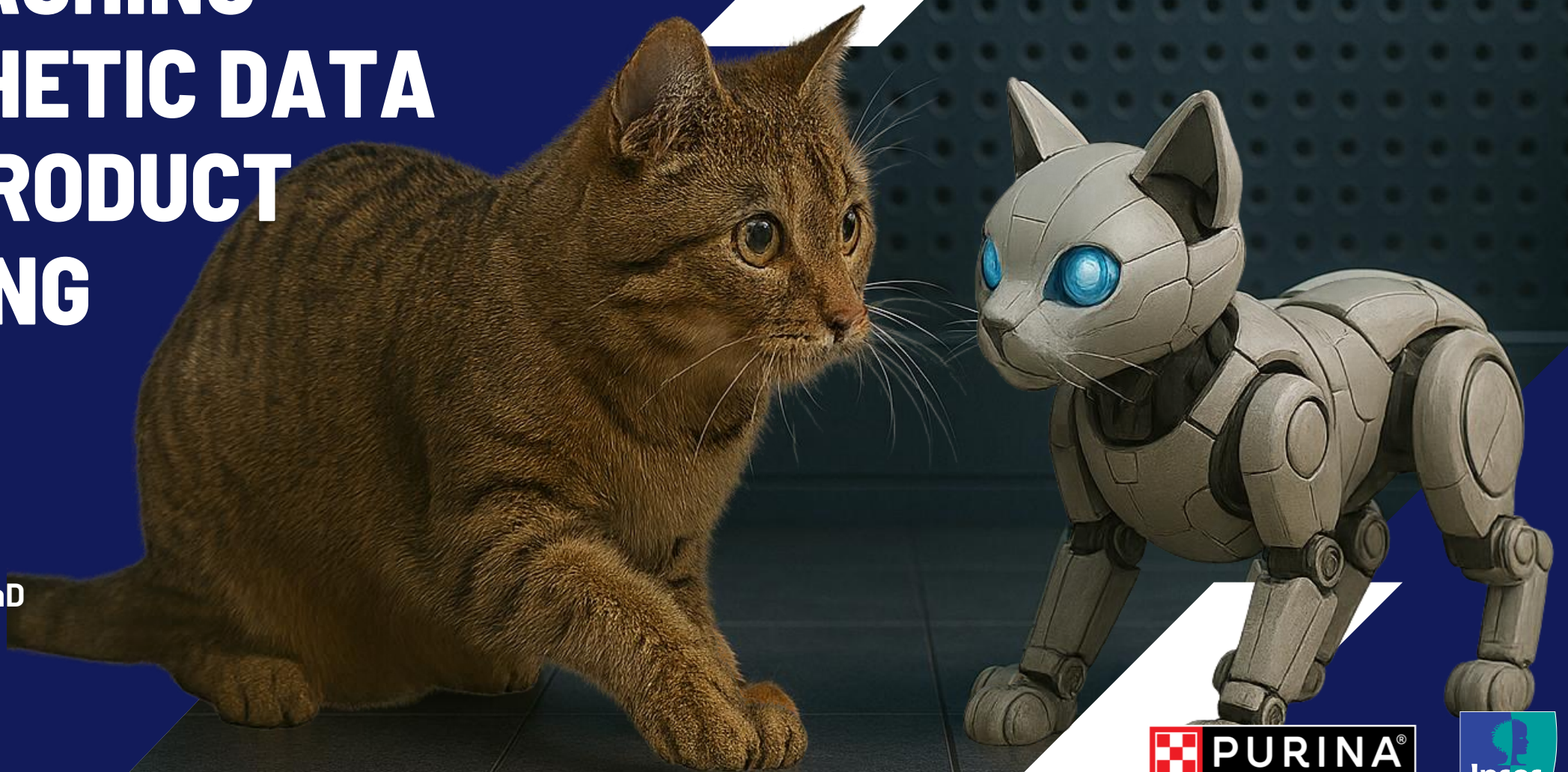
A NEW BREED OF INSIGHTS: UNLEASHING SYNTHETIC DATA FOR PRODUCT TESTING

Michael Shapiro

Senior Director
Nestle Purina

Greg Halperin, PhD

SVP, Team Lead
Ipsos Innovation



UNLEASHING SYNTHETIC DATA FOR PRODUCT TESTING



Michael Shapiro

Senior Director
Nestle Purina



Greg Halperin, PhD

SVP, Team Lead
Ipsos Innovation

1

Human Intelligence + AI

2

Synthetic Data Explained

3

Purina Case Study

4

Key Takeaways

2025 BRINGS A PERFECT STORM FOR PRODUCT TESTING





**The power of AI
depends entirely
on the quality of
its data.**

YOUR **consumers'** **experiences** **are nuanced** **and intricate**

We are all uniquely human.
Our decisions are complex,
emotional, contextual, and
often irrational.

What
is she
expecting?

What
is she
feeling?

What
is she
sensing?

What is her
environment?

What
lifestyle
brought her
here?



What is synthetic data?

Synthetic data is AI-generated information that **captures statistical patterns of real behavior.**

Why incorporate it?

- **Faster** Product Testing
- **Richer** Consumer Insights
- **Higher** Cost Savings

We focus on **data boosting** as the most practical path for product testing, paving the way for precise results

Data Augmentation



Data Boosting

Enhancing datasets with synthetic data to create a more comprehensive sample, while maintaining statistical integrity



Data Imputation

Filling in missing or incomplete data points using the available information

Synthetic Participant Research



Gen AI Agents and Persona Bots

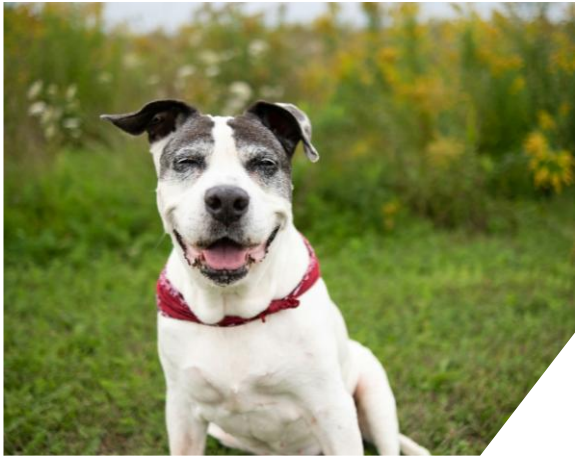
Directly engaging with the data from your studies to boost & democratize its use



Synthetic Panelists (Digital Twins)

Developing panels of 100% synthetic twins of specific respondents

Strategic Sampling Makes the Difference



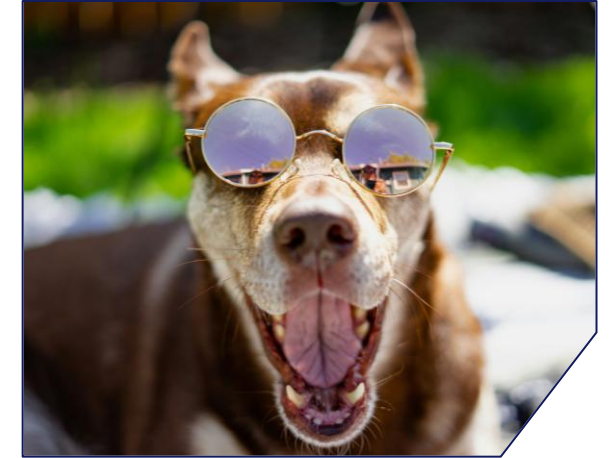
Demographics
that reflect your market



Behavioral patterns
relevant to your category



Usage occasions
that drive decisions



Attitude diversity
across mindsets

THE PRODUCT EXPERIENCE IS UNIQUELY HUMAN

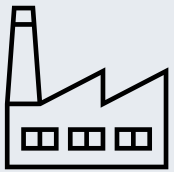
Our experiences are nuanced and intricate. Real people are **complex**, evaluate product with their **five senses**, in **context** and with **emotion**.



**THE GOAL IS NOT TO
REPLACE HUMAN
FEEDBACK, BUT TO
ENHANCE IT**

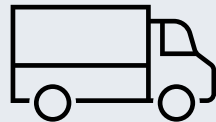
Product testing is naturally resource intensive

Time + Money + Coordination = Missed Opportunities



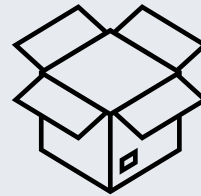
Manufacturing

Developing prototypes or products for tests



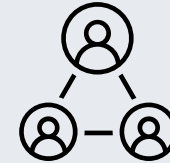
Shipping

Cost to ship products



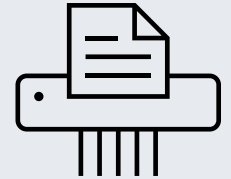
Returns

Returning used products or empty packaging



Samples

Testing with highly targeted groups



Security

Destroying test materials to protect sensitive information

The Synthetic Advantage: Speed Without Sacrifice

With the right approach, and for the right use case, synthetic data can help brands:



GO FASTER

Reduce fieldwork timing
by upwards of

50%



GO FURTHER

Unlock deeper insights
through subgroup analysis

100%



GO SMARTER

Drive cost-savings in
production and testing

20-60%

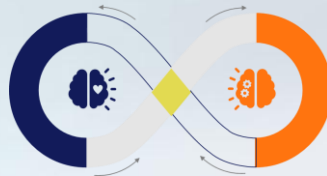
AT IPSOS, WE BELIEVE SYNTHETIC DATA WILL REVOLUTIONIZE PRODUCT TESTING



Synthetic data are not simply copies of real data

It is key to remember that synthetic data **imitates** real data, it **does not duplicate** it

Seed sample for AI
(50 Humans)



Simulation

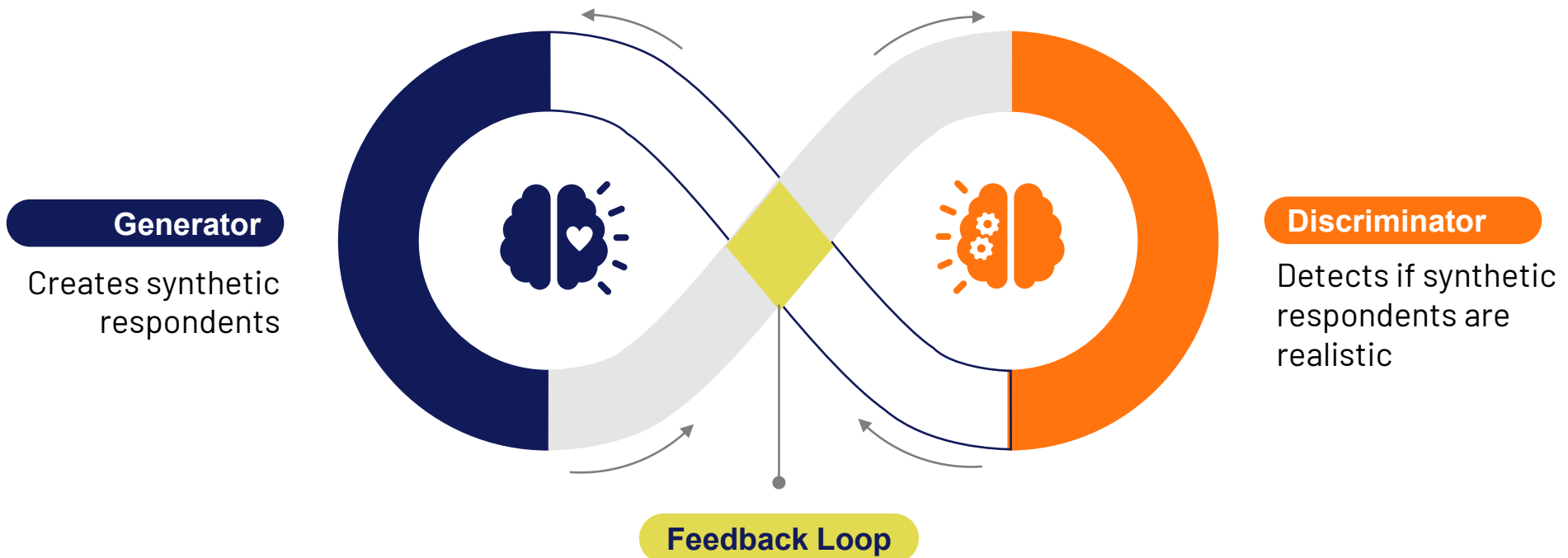
Augmented sample
(50 Humans, 150 Synths)



Synthetic data is **meant to be different** from the real data but still retain characteristics of the real data. Synthetic data, therefore, **adds variability** based on the characteristics of real data.

Inside the AI: How Synthetic Users are Made

To produce the best possible synthetic data, two neural network AI models work in tandem



A **feedback loop** continuously improves synth generation, bringing them closer to real (and trains discriminator to get better and better at detecting synths vs real)



Validation Through Rigorous Testing

Relying on previously-collected data, we conducted side-by-side comparisons of this Synthetic Users approach with traditional methods...

ACROSS MARKETS

Geographical diversity

Across Africa, Asia, Europe, Latin America, North America

ACROSS CATEGORIES

100+ Products

Across categories, including beverages, foods, personal care, homecare, QSR and pet

ACROSS METHODS

Array of objectives

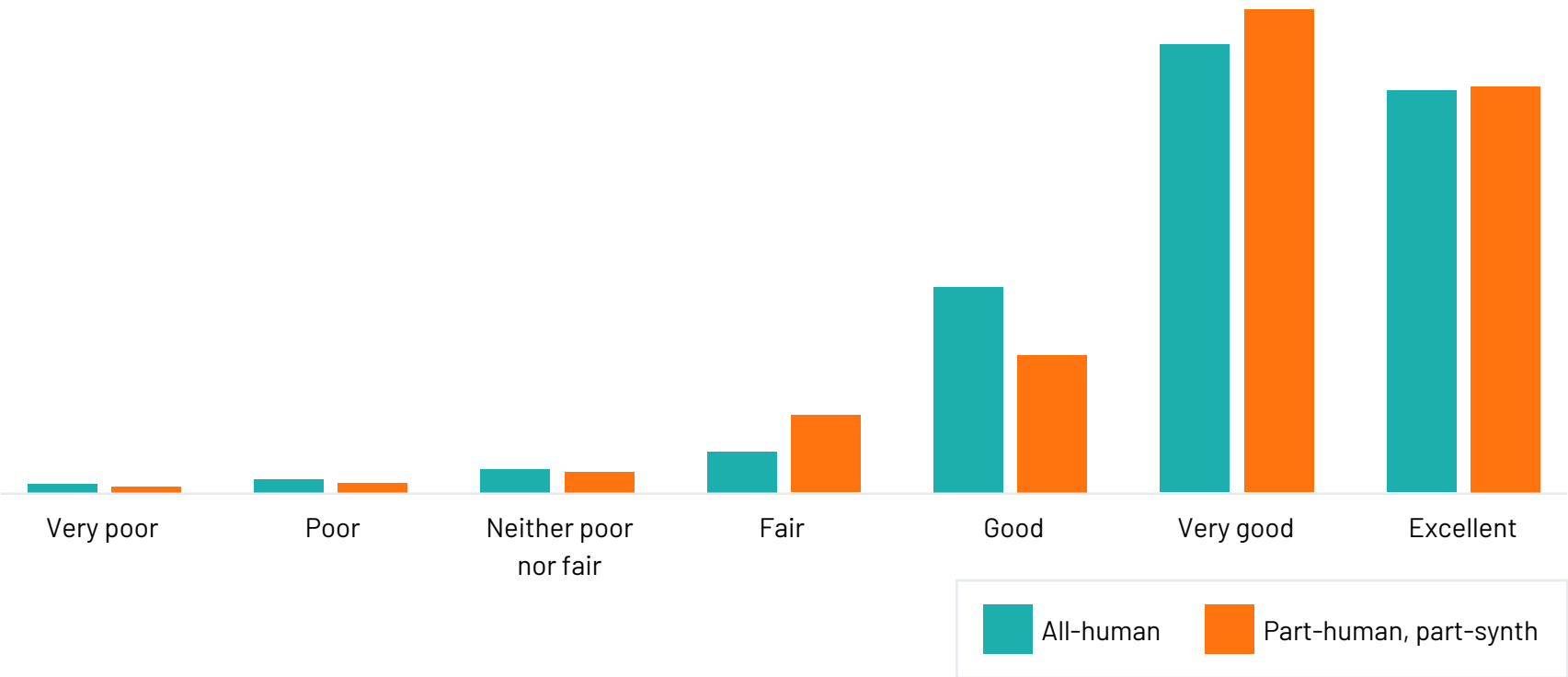
Blind and Branded
Monadic and Sequential
Complete block and incomplete block

Validations of hybrid versus real-world data show impressive accuracy in overall ratings

On average, the correlation is

0.92

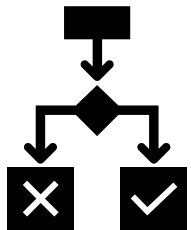
Overall Liking Scores – All Human Study vs. Human Augmented Study



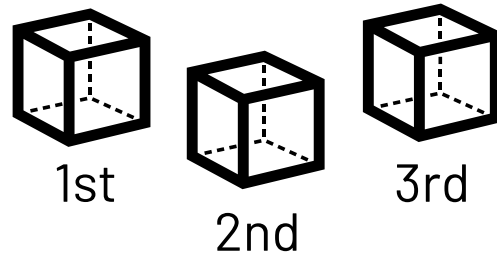
Going beyond overall ratings, we see other important consistencies



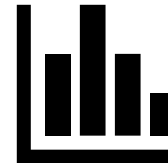
**Business
Decisions &
Action Standards**



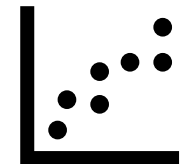
**Product Rankings
and Significant
Differences**



Data Distribution



Correlations



**But is the same
true when it
comes to our
pets?**



PRODUCT TESTING FOR FAMILY MEMBERS WHO CAN'T SPEAK



Purina's Testing Reality:

Complicated. Costly. Logistically complex.

- Gradual diet transitions required
- Extended observation periods necessary
- Complex logistics coordination
- Often hard-to-reach targets

The Validation Experiment

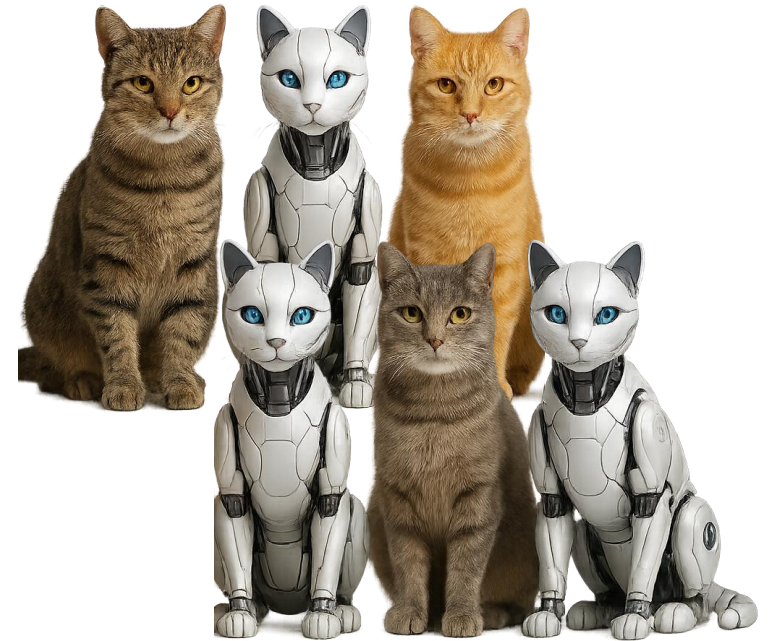
By comparing our “hybrid” data back to the original data set, we could see if the outcomes would have been the same



n50 Real Cat Owners



n150 Synthetic Cat Owners



n200 Hybrid Cat Owners

KPIs for the hybrid consumers were nearly identical to the original results for all real consumers.

Looking across the total sample as well as subgroups, we see no statistical difference in outcomes

	Cat Owners		Wet Food Buyers		Brand Buyers	
	Original Real Humans (n=225)	Real + Synth (n=50+150)	Original Real Humans (N=190)	Real + Synth (n=50+150)	Original Real Humans (n=174)	Real + Synth (n=50+150)
Purchase Intent	=		=		=	
Overall Liking- Owner	=		=		=	
Overall Liking- Cat	=		=		=	
Overall Appearance	=		=		=	
Comparison to Expectation	=		=		=	
Has a taste my cat will love	=		=		=	
Is made with high quality ingredients	=		=		=	
Fits with the brand	=		=		=	
Would help me show how much I love my cat	=		=		=	
Would create excitement and eagerness for my cat	=		=		=	

In addition, diagnostic results remained consistent, identifying the same key optimization opportunities.

The same factors appeared as areas of improvement, regardless of method.

	Cat Owners		Wet Food Buyers		Brand Buyers	
	Original Real Humans (n=225)	Real + Synth (n=50+150)	Original Real Humans (N=190)	Real + Synth (n=50+150)	Original Real Humans (n=174)	Real + Synth (n=50+150)
Liking: Overall Taste						
Liking: Taste KPI #1						
Liking: Taste KPI #2						
Liking: Overall Texture						
Liking: Texture KPI #1						
Liking: Texture KPI #2						



Business Impact for Purina

Adoption of a synthetic approach would allow:

- Enhanced subgroup insights
- Optimized resource allocation
- Accelerated innovation process

The Key Takeaways?



Synthetic data works,
reaching the same business
decision in 95% of cases.



**Sample quality determines
everything.**



**Augmentation over
replacement** for product
testing.

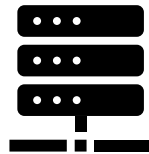
But there are also considerations...



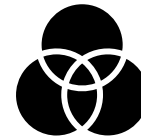
**Relevant Training
Data is Crucial**



**Data Boosting, Not
Replication**



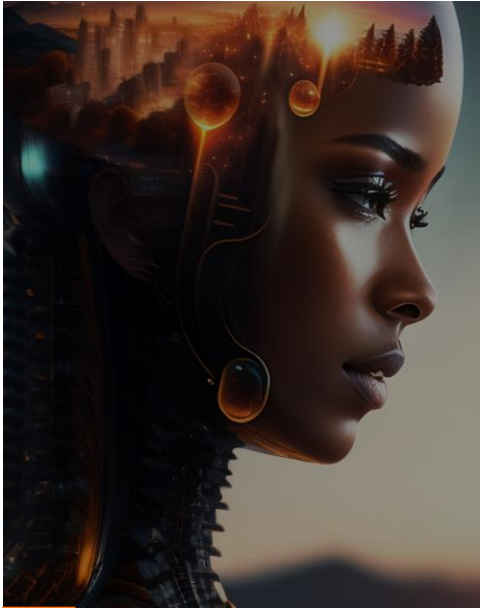
**Validate, Validate,
Validate**



**Risk Assessment
is Essential**



Purina is committed to exploring GenAI research methods



AI- Powered Innovation Process: INNOEXPLORER

Leverage generative AI to innovate winning concepts from unmet needs discovery through concept generation, screening and final validation.



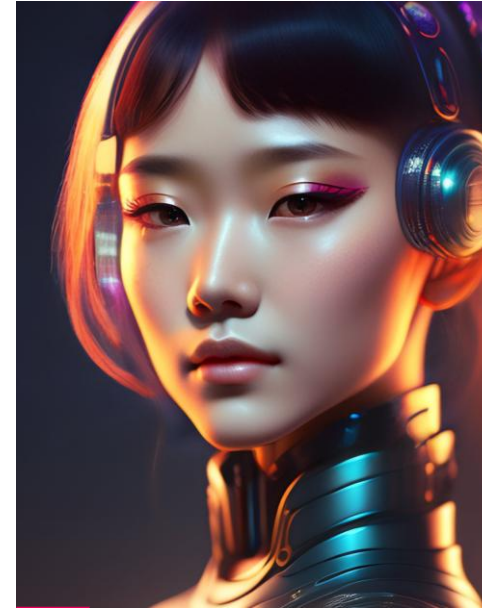
InnoPredict AI

Enable business teams to replace some of their traditional idea screening to reduce costs and hasten bringing new innovations to market.



Segmentation Persona Bot

Leverage Generative AI to assist with bringing to life Purina persona bots to build empathy with Marketing team and enable us to interact to ask questions and experiment with creating new commercial and product ideas.



AI-Assisted Online Community

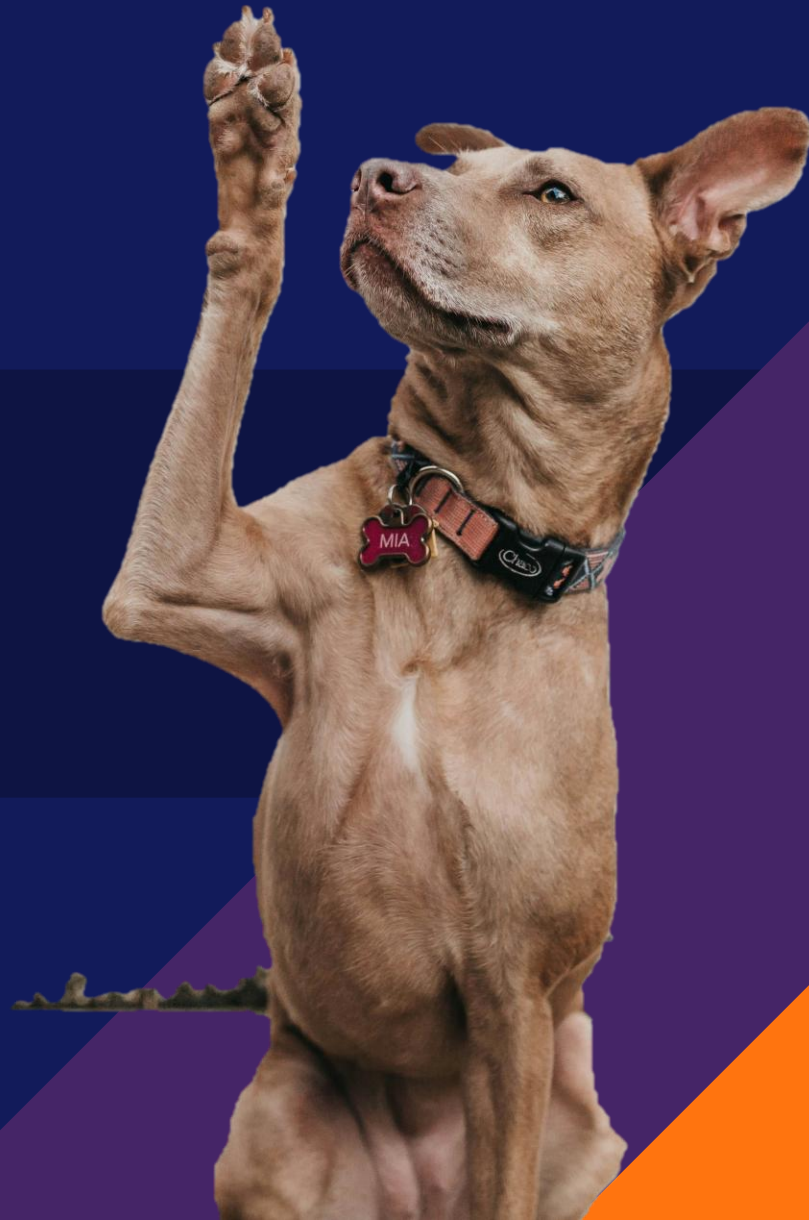
GenAI to assist in deciphering deep meaning and themes in Purina product tests executed on the Ipsos Community Platform to analyze qualitative and video responses for enhanced storytelling from "the pet's POV."



GenAI-Powered Transcript & Knowledge Curation

Developed Ipsos Facto Orra platform. Use of AI to summarize existing transcripts and other research data to add greater return on research investment.

QUESTIONS?



THANK YOU

Unleashing Synthetic Data for
Product Testing