Future of Strategy: Pressure testing synthetic research

Chris Mele, Hugo Alves and Tamar Simpson
Source: WARC Best Practice, December 2024

Putting peer-reviewed, in-market GenAI synthetic research tools to the test by stacking process, inputs/outputs, insights, and directional strategy against traditional organic methodology.

- A side-by-side comparison of human-to-human research and synthetic research using the same briefs for the research lead and the robots.
- The test compared a national market research program that combined qualitative and quantitative across 20 geographical markets, three unique audience segments, and a spectrum of demo and psychographic.
- Interview insights were then run through a stack of GenAI tools to see how they were at crafting strategy and visual identity.

Why it matters

Strategists, agencies and marketers have a slew of new AI-based tools to employ and the promise is that these tools will increase efficiency. This test compares the output of a traditional, human-led research project with the results from an AI-based tool.

Takeaways

The side-by-side experiment was evaluated looking at several factors: The time it took to get interviews completed; the quality and depth of the interviews; the insights; and pulling insights through to direction.

- **Time for interviews**: Synthetic research was very fast, taking a day to complete interviews compared to the two weeks it took our research lead.

- **Quality and depth of interviews**: Synthetic research interviews were text only, because they weren't done with real people, so not fully comparable to the human-to-human interviews. But comparing text to text, the synthetic research held its own.
- **The insights**: Synthetic insights summaries may read a touch clumsy, but the reality is that they aren't so far removed from where we landed with analog.
- **Pulling insights through to direction**: We employed ChatGPT (which is good at strategic, logic-based writing), Claude (which is good with creative writing), and Ideogram (which is ok at creating visuals), and Google's NotebookLM for audio.

If resources like time and money are a major factor, consider exploring synthetic research to get to some quick and dirty insights (or test concepts, run scenarios, etc). But if working in a highly nuanced or competitive space, or internal deliberation and alignment are a major concern, this technology probably isn't ready yet for that.

Intro

With a near-constant stream of hype around the immediacy for marketers to shift to AI-based solutions across the CX lifecycle, it's becoming increasingly difficult to separate white papers and hype demos from "here today", turnkey technologies that can be implemented safely and efficiently to improve workflows, outputs, and ultimately, business outcomes.

Those AI-based solutions include synthetic research, which is a process that leverages large datasets and natural language processing to arrive at research insights and strategic output, as opposed to human-to-human research (ie. interviews, surveys, etc). It seeks to improve on or enhance traditional research by both increasing efficiency (scripts can produce in minutes results that may previously have taken days or even months), and immensity (limited only by the scope of scrapable content on the internet).

To turn theory to practice, and to gain some hands-on experience and first-person perspective, we cracked open an in-progress rebrand and website relaunch with The James Beard Foundation. The James Beard Foundation is both a non-profit and a major driver of American food culture. It is thus a nuanced study of audience perception and relationships. We compared a national market research program that combined qualitative and quantitative across 20 geographical markets, three completely unique audience segments, and a spectrum of demo and psychographics, to a synthetic process. We briefed the systems at **Synthetic Users** with the same directives our analysts received to design, conduct and synthesize the program.

As a bonus, following the "execution" of the research and insights gathering, we took it one step further and leveraged additional best-in-class GenAI platforms to move through a sketch of the strategic and creative process of rewriting our brand strategy and generating a new visual identity direction.

Setting up the experiment

To ensure we were setting up the processes and prompts properly, we enlisted the help of Synthetic Users co-founder and Chief Product Officer, **Hugo Alves**, who graciously offered to help guide the process on the platform, mitigating the potential for user error or incompetence.

The robots received the same brief that our research lead did. We gave them market geos and

psychographic/demographic splits via our recruiting screener, and even the discussion guides that we used to conduct the interviews. It's important to note that the Synthetic Users system relies on AI agents, and not just question-and-answer language processing and results generation. By that, we mean that the AI was given agency to conduct the interview as it saw fit. Much like a human researcher, if it heard something interesting, it could ask follow-up questions or press the interviewee to expand on a topic or explain itself more deeply.

Once the interviews were complete, we also asked our agents to summarize their findings in an insights summary. What were the topline takeaways? Were there any surprising insights? What about any hard-hitting verbatims that might serve well as a pull quote in an executive summary?

Finally, we added a couple of final tools to the stack to pull the project simulation through to strategy and creative output.

Running the process

Rather than directly translating the complete analog interview guides into one-to-one synthetic conversations, we took a more strategic approach that allowed us to grant the system's agents more autonomy to do their own "thinking", and produce more nuanced and thorough results.

Our prompt engineer identified the thematic blocks within the interview question sets – addressing everything from culinary heritage to sustainability practices – and used those themes as pillars to inform a roadmap for automated research goal decomposition.

Each thematic block was manually transformed into a high-level research goal, and those goals were automatically broken down into full-cycle, targeted studies. The system then leveraged large language models to generate detailed profiles for each user group based on the provided demographic and psychographic criteria, ensuring our synthetic users would authentically represent JBF's diverse audience segments while maintaining consistent and realistic perspectives across different interview contexts.

The system generated two kinds of outputs. The first was raw interview transcripts for each user group – providing direct, unfiltered synthetic user responses that read much like a transcription from an analog verbal interview.

The second was an automatically generating comprehensive summary for each thematic section across all user groups. These summaries distilled key findings, patterns, and insights that emerged, providing an efficient way to compare perspectives across different audience segments while maintaining the depth and nuance of individual responses. We essentially produced a like-for-like research findings report, very similar to the presentation that the analog process produced.

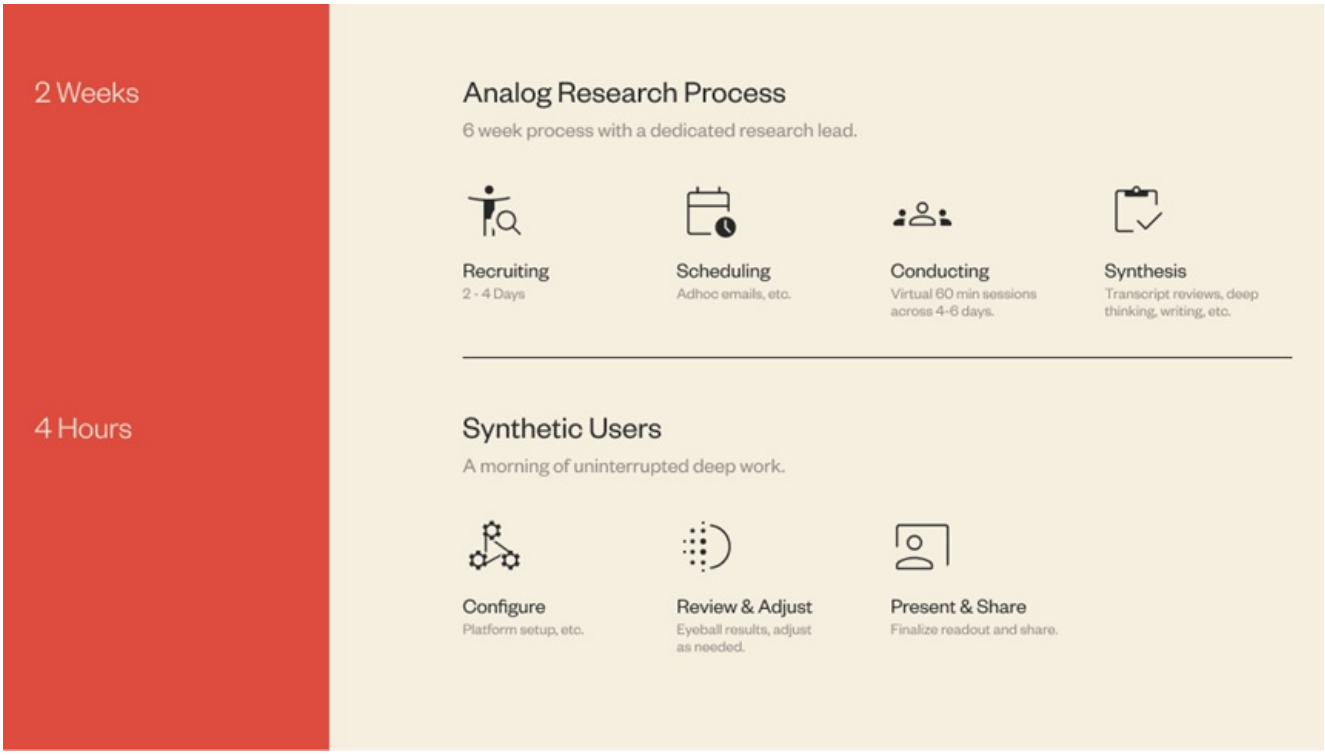
The results

All in all, the platform dutifully produced a like-for-like output of qualitative research raw data and top-line summary. Fast. Very fast.

Let's take a look at the results and dig into a few key assessment factors.

Time to conclusion

What played out in our analog process as a two-week process of recruiting, scheduling, conducting, and synthesizing was completed in a single morning. The time invested by our prompt engineer to understand the assignment was a wash when considering the onboarding of a new analog team member to help conduct research, so we did not factor that into our comparison. We ran 16 physical interviews during that two-week process (combined with a quant study that garnered 2,500 responses), and Synthetic Users completed a total of an eye-catching 160. Should we have decided to scale the program by expanding the coverage to global markets, for instance, the return on time invested would really start to widen the gap and tilt in favor of Synthetic.



Winner: Synthetic Research

When purely looking through the lens of speed from “go” to raw research completed, there is no argument to be made against the synthetic approach. It's faster, and more efficient.

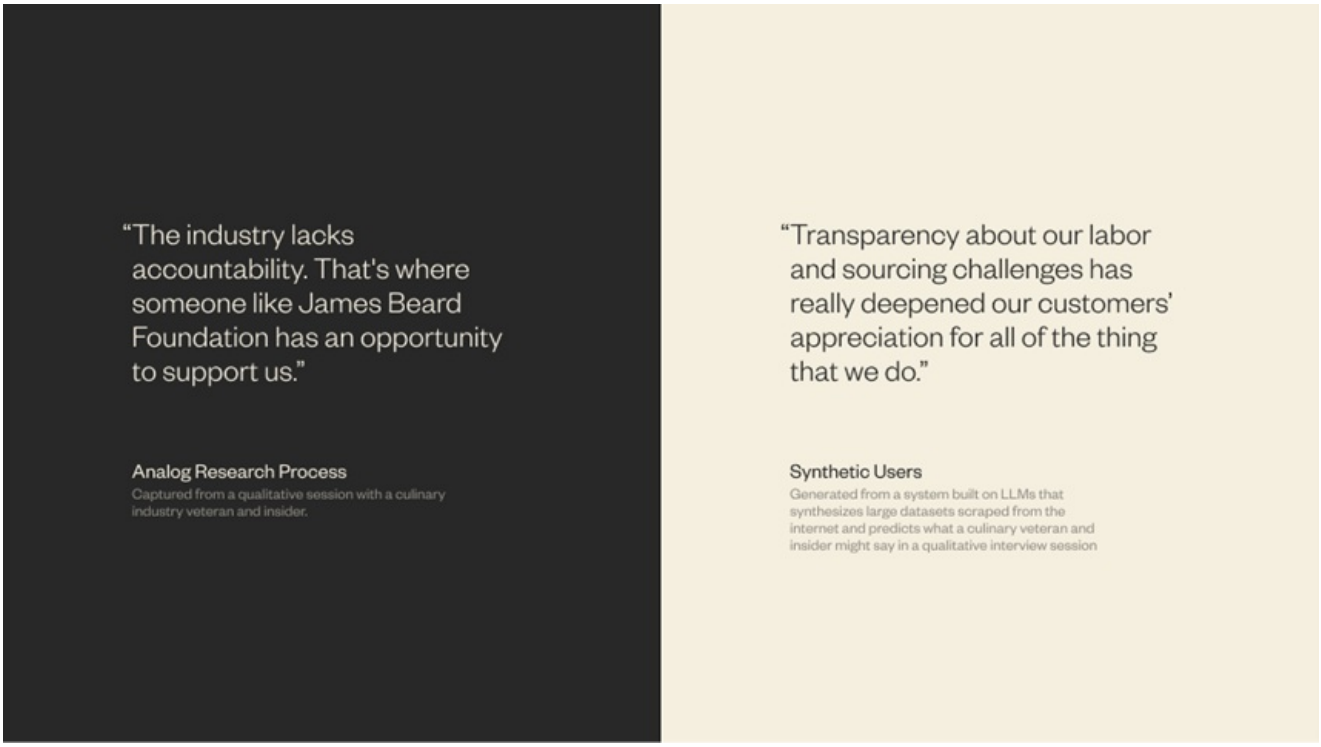
Quality & depth of interview

The interviews themselves exist only in text form. Because they aren’t real. This presents a slight conundrum as a qual purist, as I believe that empathy and understanding are built by the relationships we build with our subjects throughout our sessions. We are part therapists, part journalists, and part documentarians, and the true insights come from the awkward pauses, tonal variations, and tightened lips that relate a deeper tension or excitement that lives *below* our words. That said, if you were purely looking at the text of the interviews themselves… they were pretty solid! Yes, there were some expected-if-not-trite responses (**“Seeing the positive impact of sustainable practices firsthand has been incredibly motivating…”**), but frankly, we get a lot of that in our analog work, and parsing through it is part of the challenge that is producing valuable insights. I can even point to a few verbatim nuggets that put a fine point on a few of our own key takeaways that I think would have played well in our executive board summary.

I'd also be remiss to not take a moment to congratulate AI for not skipping a moment to big up itself:

"One tech-driven solution that has really impressed me is the use of AI-powered inventory management systems." - Noah Patel

Oh, were you impressed, Mr. Patel?

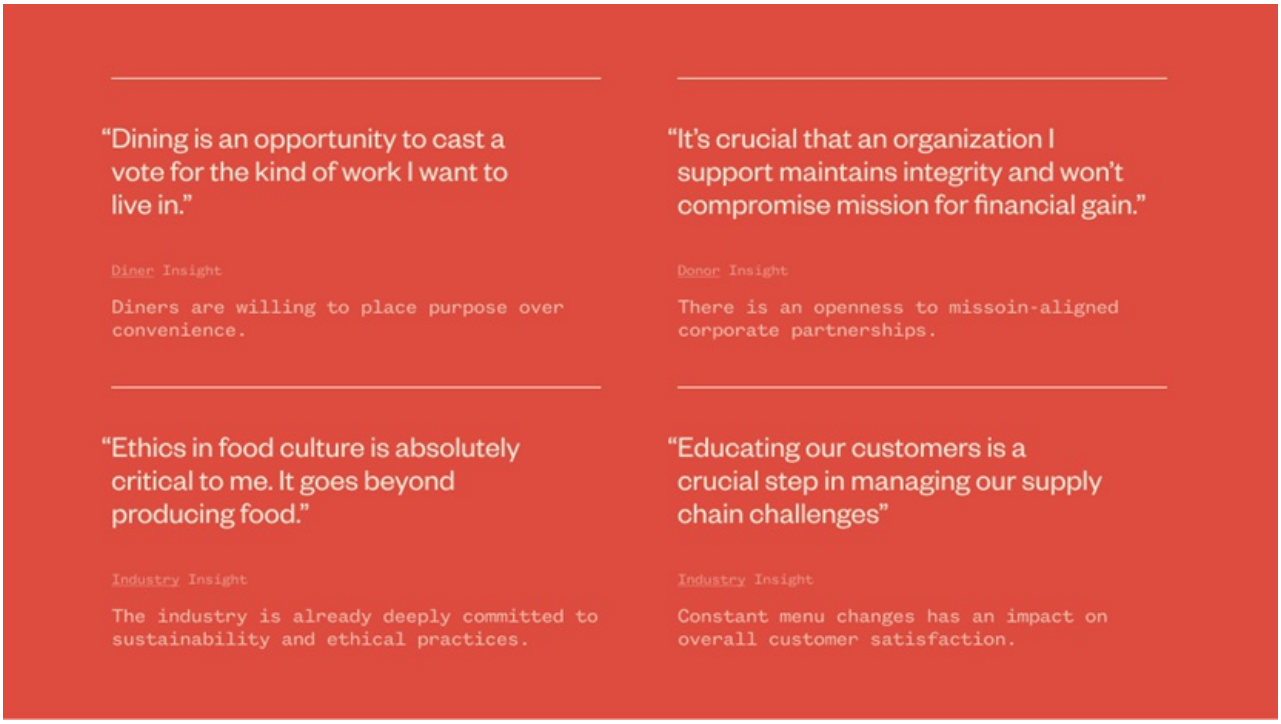


Winner: Analog Research (but it was closer than expected!)

When push comes to shove, the human perception element wins out. Words alone simply can not replicate the deeper empathy and understanding that we can gain from body language and human interaction (more on this later).

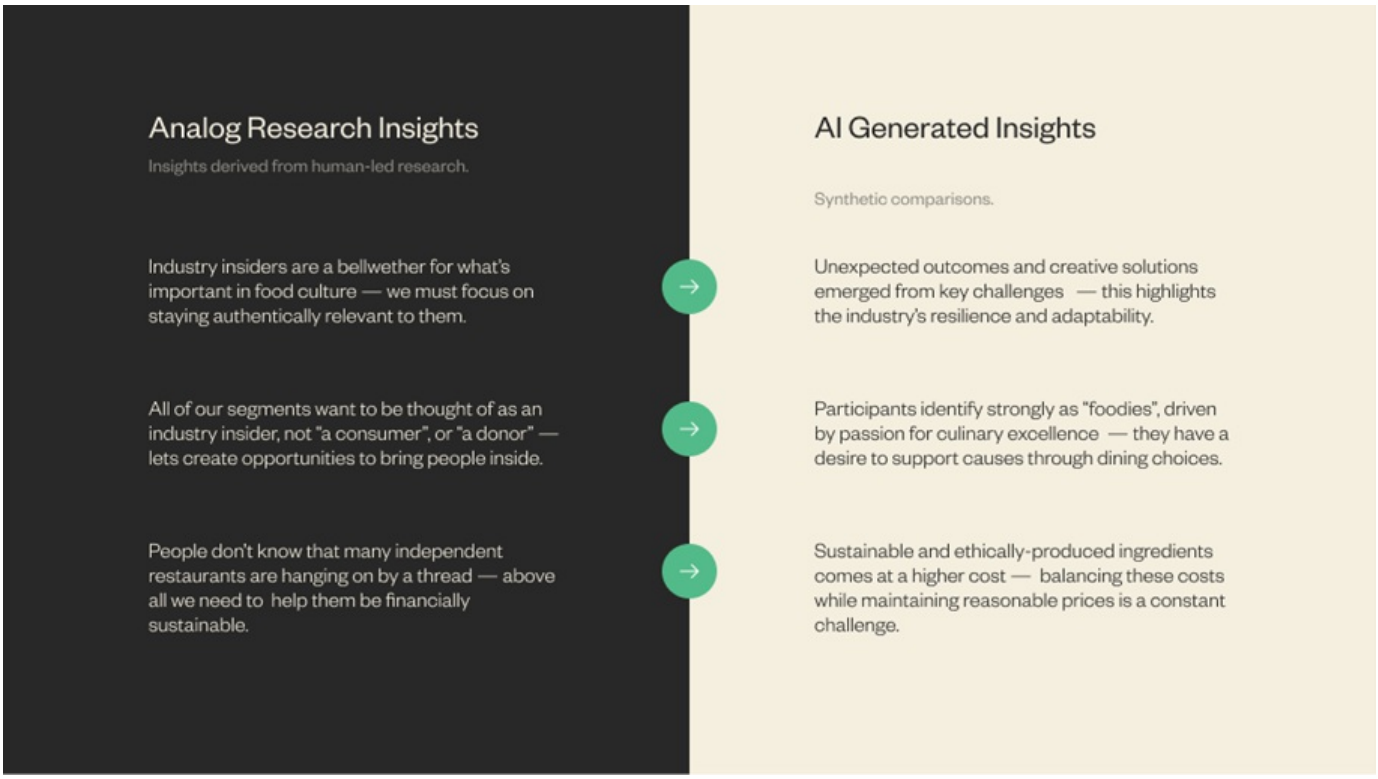
Insights

The insights produced *mostly* read as slightly robotic. They weren't *wrong* per se, but the overall body of work felt underseasoned and would have been difficult to work with as a spark to create a strategy. I would maybe compare the average insight output to that of a junior-to-mid-level strategist. That said, there were a few interesting points raised, and we absolutely LOVED the "surprising insight" category, which seemed to add a more human-like quality to the outputs and actually outlined some intriguing direction that at the very least would have served as fodder for deep debate amongst the JBF leadership team.



Winner: Synthetic Research

AI is good at parsing large data sets and summarizing commonalities. The insights summaries we received from Synthetic User may read a touch clumsy, but the reality is that they aren't so far removed from where we landed with analog. With the right amount of tuning, I believe we may have been able to reach parity, and in the interest of provocation, the tie is going to the challenger.



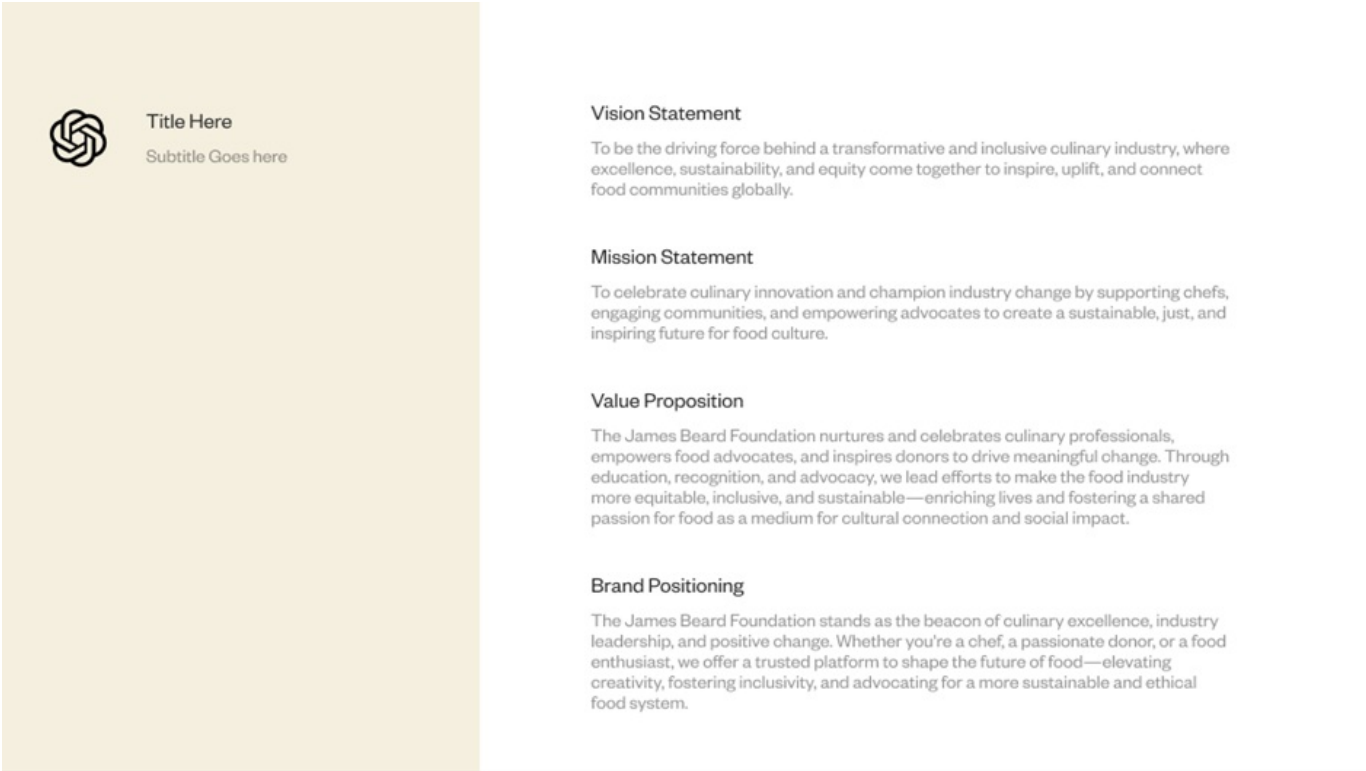
Pulling insights through to direction

This is where it starts to get interesting. The Synthetic Users platform is focused on the specific stages of the value chain where the directive to produce insights becomes activated, and raw, summarized insights are returned for synthesis and direction.

But being this is a piece on the future of strategy, we were curious to go as far as possible into the synthetic process, so we enlisted *Synthetic User*'s friends ChatGPT (which is good at strategic, logic-based writing), Claude (which is good with *creative* writing), and Ideogram (which is *ok* at creating visuals – cut it some slack, it's a robot), and finally, Google's NotebookLM for some audio fun.

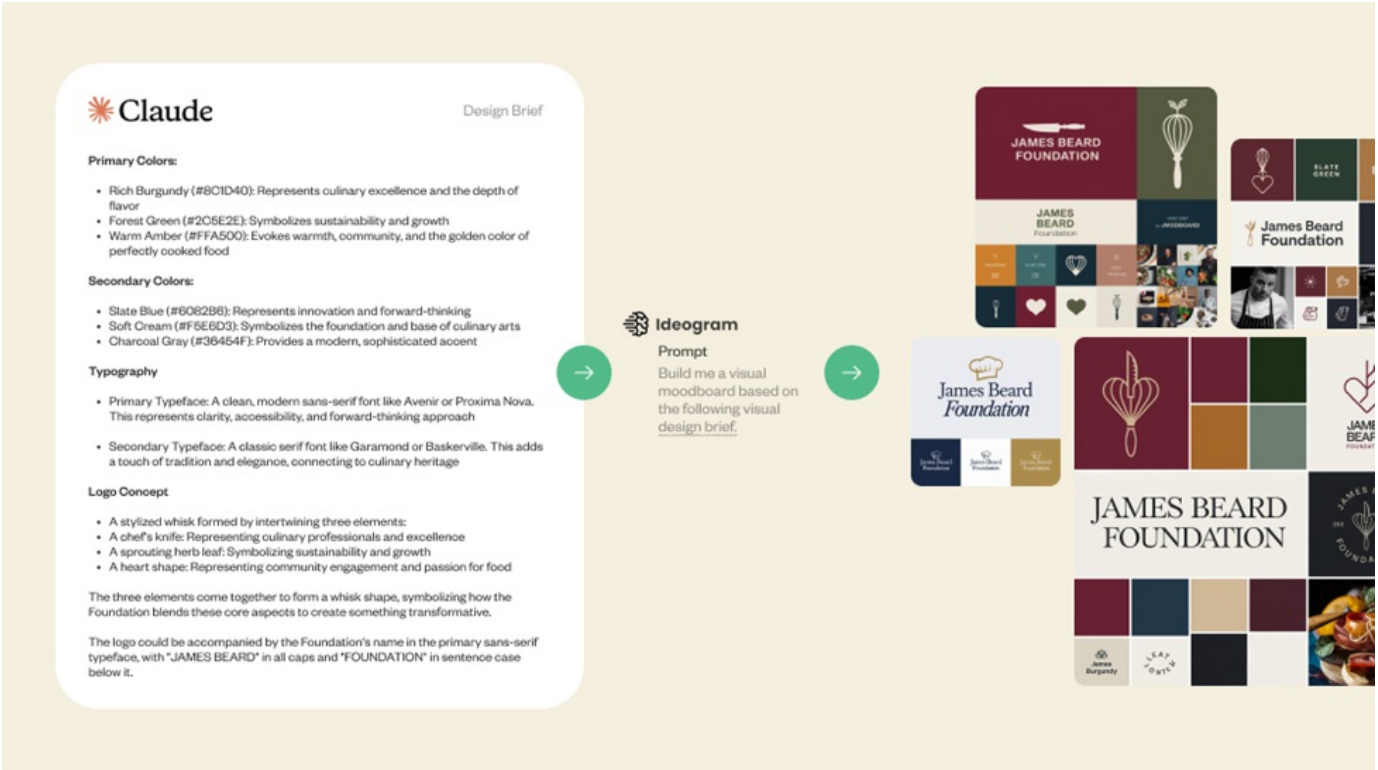
The results were fascinating. Here's a quick-fire summary:

- ChatGPT – we fed all the raw synthetic interviews and insights summaries into ChatGPT, as well as the analog "brand house" framework we used to set our new positioning and strategy (not the content itself, just the elements themselves, their descriptions, and their business purposes), and asked it to write its own.
    - Result: It lacked a bit of the nuance our own strategy possessed, but it was shockingly passable.
    - Grade: A – I was personally pretty blown away by this result.



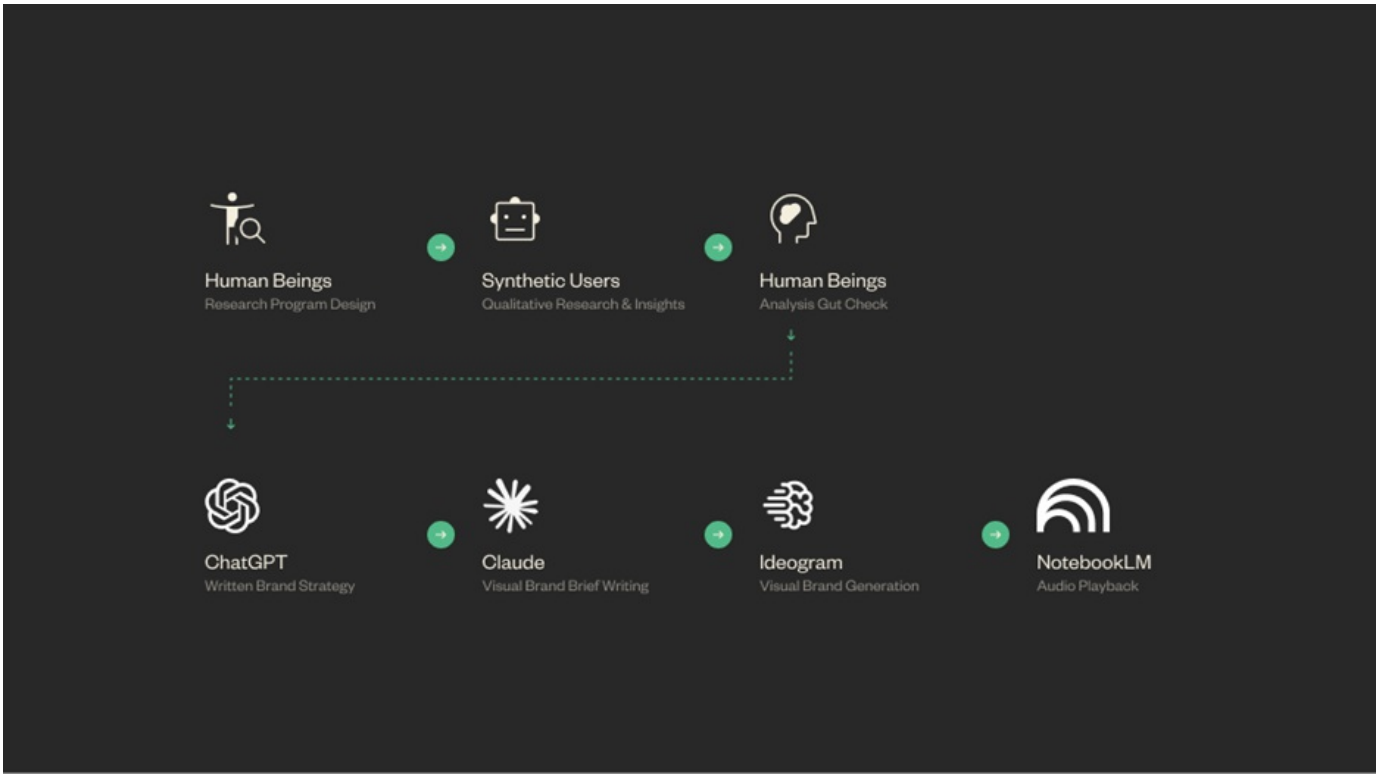
- Claude – We then passed ChatGPT's brand strategy on to Claude and asked it to write a visual identity brief for a rebrand of James Beard Foundation, complete with color palette, typeface selection, written logo concept and general design strategy.
    - Result: The color and type choices were elegant and refined, and not wholly dissimilar from some of the early exploration work our visual design team completed during the pitch. The logo concepts, however, were completely hamfisted, calling for a "stylized whisk" with an intertwining chef's knife, herb leaf, and a heart to represent community. For shame, Claude. Go back to design school!
    - Grade: B – you almost had it until we got into the more conceptual space…

- Ideogram – finally, we tapped Ideogram for its ability to “generate realistic images, posters, logos and more”. We expected the system to work off Claude’s clearly written direction to create on-brief visuals that could serve as the basis for a 360 visual identity package that would have to be completed by a human. We asked for a set of general mood boards to include an evolved wordmark and general visual treatment. This is really where the quality gap started to widen in my opinion.
    - Result: Rendering the color palette and type direction is a simple task and it executed well enough, but again, once we moved into the more creatively nuanced territories, things fell apart. I’ll once again caveat that generative brand visuals are not an area where I consider myself proficient, so it could be that with the right amount of fine-tuning, we could get to something passable. That said, if a professional designer brought this work into a design crit, it would be a challenging conversation to have.
    - Grade: F – sorry bud.



- NotebookLM – for extra credit, we also fed our raw insights output into Google’s NotebookLM and used their Audio Overview feature to create a conversational 1v1 podcast of two “experts” discussing the results. Aside from the general AI novelty, I’m not sure about the use case, but it was fun to listen to! It sounded like two average LinkedIn thought leaders with radio DJ voices conducting a bland but on-brief podcast episode, all the way down to the ums and uhs. I also laughed out loud at “It’s not all sunshine and farm fresh ingredients”. We only occasionally entered the uncanny valley and I had a moment of delightful disbelief that humanity has pulled off this particular magic trick.

To recap, below is the end-to-end workflow and tooling we used to arrive at our “rebrand”.



Putting our work under client review

I thought it would be interesting to put our results through “client review”, and so we enlisted James Beard Foundation’s VP, Marketing and Communications, Tamar Simpson to review our results at each stage of the process and weigh in. Here’s what she had to say:

> “While the synthetic insights were surprisingly reasonable, and the reports fun to read, the process required a partnership that revolved around discussion, feedback and iteration to fully trust the outcomes. The human connection is what made our work so valuable.”

Tamar Simpson
VP, Marketing and Communications
James Beard Foundation

Summary

So what did we learn?

Efficient, blazing-fast data analysis is indeed the sweet spot.

If resources like time and money are a major factor in your research, you might consider exploring this route to shake up some quick and dirty insights (or test concepts, run scenarios, etc).

If you are working in a highly nuanced or competitive space, or internal deliberation and alignment are a major concern, this technology probably isn't show-ready for you – though as an IC you might consider some sidebar workflows to help you get your gears turning.

Scale is also an important lens to apply. The larger and wider the research data sets become, the more difficult the task for a human to find those trendlines and make sense of them (essentially, this is where qual becomes quant, more finely tuned around multiple choice sets). There's a continuum of surface area to the investment case for rigorous, broad research, and I can imagine scenarios for SMBs and Fortune 100s alike where this will become a viable layer in your insights stack.

Finally, let's all keep in mind how nascent these applications of predictive and generative systems are. When you step outside of yourself, if you aren't at least a little bit impressed by some of what we produced here, I tip my cap to your (word that means you are hard to impress).

The future is going to be wild.

About the authors

Chris Mele
Managing Partner, Siberio.io

Hugo Alves
Co-Founder and Chief Product Officer, Synthetic Users

Tamar Simpson
VP Marketing & Communications, James Beard Foundation

Ascential Events (Europe) Limited
Americas: 229 West 43rd Street, 7th Floor, New York, NY 10036, United States - Tel: +1 212 201 2800
APAC: 5 Shenton Way, UIC Building #10-01, Singapore 068808 - Tel: +65 3157 6200
EMEA: 33 Kingsway, London, WC2B 6UF, United Kingdom - Tel: +44 (0)20 7467 8100

www.warc.com

WARC