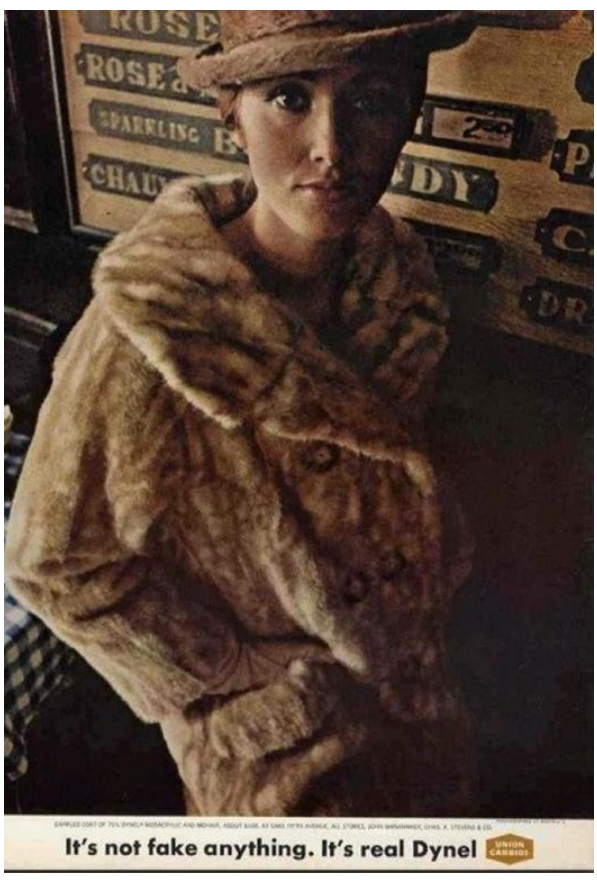
McDonald Minute

Spending a little time on big ideas

Summer 2025

Keeping It Real: Authentic Uses for Synthetic Data



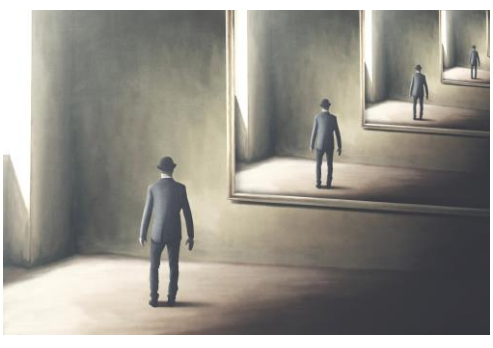
If you're a Baby Boomer like me, you were raised in a mid-century world that fetishized authenticity and apologized profusely for anything faux. Changes in cultural norms have elevated vegan leather but in the 1960's and 70's, plastic imitations of anything were scorned as cheap and cheesy. Back then, new polymer-based synthetics had to be cleverly positioned to make them acceptable: "It's not fake anything. It's real Dynel." Even products we'd *grown up with* were pressure-tested for authenticity, like Coke—repositioned as "the real thing" in 1969, a watershed year of living earnestly. And *people* needed to be real too. Holden Caulfield, hero of the anthemic book of our generation, spent most of his time outing "phonies." (Bonus points for Millennial readers who said *Catcher in the Rye* without asking Chat!)

From Sincere to Synthetic

We can trace our ongoing pursuit of "brand authenticity" – which is a concept we still struggle to define and measure – back to that mid-century emphasis on being earnest and unfiltered. But 20th century authenticity has evidently run its course, overwhelmed by the more powerful forces of 21st century technology and the strangling, dehumanizing embrace of the artificial. We're now a culture where people look to chatbots for mentorship and even friendship—and where most of us would struggle to recognize a deep fake even of ourselves.

Signaling that same trend in our own field is the buzz about "synthetic data" – a term bandied about but largely misunderstood. Research forums on synthetic data leave most people scratching their heads but still at the edge of their seats because those *very words* hold the thrill of alchemy. We are, all of us, fighting the suspicion that fake might somehow be *better than real*.

To be clear, there are several well-established use cases for synthetic data. Two of the most compelling involve very large, fully anonymized datasets that can effectively model outcomes or be used simply to train models. Synthetic health datasets, for instance, have the benefit of enormous scale and a process of parameterization so thorough that they may, indeed, be every bit as good as the real thing. Not so apparent, however, is what else *market research* alchemists can safely and cost-effectively do with synthetic data. Virtually every use case I've heard discussed – from AI-created personas of market segments to synthetic respondents for survey pilots – rely heavily on extrapolation and probabilistic invention.



The AI tools used to fabricate these synthetic samples create the *illusion* of enlarging the database without actually narrowing the confidence interval. In fact, the very idea of applying significance tests to these synthetic databases is suspect. Synthetic respondents contemplated for market research are byproducts of datasets too small to be statistically replicated in a way we can trust. They will be riddled with deep fakery invisible to the naked eye. A sacred mission for our industry sector is measuring change. Synthetic data is no way to do it.

This old Baby Boomer gets the vapors whenever she hears the argument that synthetic data may actually be *preferable* on grounds that the real thing is untrustworthy. ("It's not fake anything, it's real synthetic data.") The idea of trading a warty reality for a smooth and shiny fake one raises lofty questions about the meaning of "real," and we should certainly question the value of any simulacrum whenever there are doubts about the credibility of the original. But let's keep this conversation on a practical plane: the case for using synthetic respondents, based on the argument that they're faster and cheaper to source than real people, is a premise that rings hollow when you do the math and consider not just the risk but the missing sentience and insight. Real people might not always be "honest" with you, but they have a personal truth to reveal.

True to my minute mission, this is not the proper place to drill deep on complex issues. And, of course, everything about AI is a moving target so we should *almost* never say never. But this is a gotta-get-it-right moment for the industry, to avoid losing our footing and our credibility in a way we can't easily recover. There's already too much to do, absorbing AI into our basic processes, without distractions. We need to focus on the highest priorities and the biggest returns. I look forward to sharing more detailed discussions we're having at NAXION and with colleagues throughout the industry over the next few months. How we'll all navigate authenticity in a deep fake landscape remains to be seen.

About the Author

Susan Schwartz McDonald, Ph.D.
Chief Executive Officer
215.496.6850
smcdonald@naxionthinking.com

Susan's career focus has been on the development and protection of robust brands, and the research methodologies needed to support them. She has contributed to the evolution of many standard research techniques, and she writes frequently on industry topics and issues of broader interest. In deference to changing times, the snappier-paced McDonald Minute replaces Susan's long-form Smartmouth blog, but with the same goal of connecting cultural themes to business challenges. She holds MA and PhD degrees from UPenn's Annenberg School of Communication.

About NAXION

NAXION is a nimble, broadly resourced boutique that relies on advanced research methods, data integration, and sector-focused experience to guide strategic business decisions that shape the destiny of brands. Our century-long history of innovation has helped to propel the insights discipline and continues to inspire contributions to the development and effective application of emerging data science techniques. For information on what's new at NAXION and how we might help you with your marketing challenges, please visit https://www.naxionthinking.com/

© NAXION, Inc. All rights reserved.

NAXION RESEARCH › CONSULTING