



Brightseed Expands World's Largest Proprietary Bioactive Dataset to 21 Million Compounds, Strengthening AI-Native Innovation Platform

Rapid dataset growth underscores compounding data advantage and accelerates discovery across human health, microbiome, and emerging biological domains

SAN FRANCISCO, Calif. — April 15, 2026 — Brightseed, the life sciences innovation platform company powered by AI, today announced a major expansion of its proprietary bioactive dataset. Now at 21 million natural compounds, its database has nearly doubled since 2023. This growth is coupled with massive gains in efficiency as discovery costs have decreased over 98% since the company's start in 2018.

Data remains a foundational pillar of Brightseed's platform and a core driver of its differentiated AI and machine learning capabilities. Combined with advances in profiling and data analysis, this expansion increases both the breadth and depth of the dataset—further strengthening the platform's ability to generate novel, high-confidence insights.

Brightseed's dataset enables the generation of functional inference models across over 23 health territories. Its corporate partners use these to make accelerated innovation advances across agriculture, health, nutraceuticals, animal health, and beyond. For example, one of its consumer health partners is leveraging Brightseed's platform to model and discover complementary pathways in the GLP-1 signaling cascade, with a growing pipeline of candidate molecules already identified.

"The real story here is not just the size of our dataset, but its productivity," said Lee Chae, PhD, Co-founder and CEO of Brightseed. "Brightseed's platform is built off of the world's largest biologically connected and actionable library of proprietary natural compounds, and we've only tapped a small percentage of this dataset so far. There's still so much gold for us to explore and discover. These endless possibilities are our moat."

Unlike public datasets or generalized AI training data, Brightseed's platform maps its proprietary dataset of small molecules to biological mechanisms and human health outcomes, enabling discovery and validation at unprecedented scale.



This dataset powers Brightseed's innovation platform, including its AI discovery engine, Forager™, enabling customers to:

- Identify relationships among novel bioactives and mechanisms not found in public literature
- Evaluate biological relevance and feasibility earlier in development
- Generate differentiated, IP-ready innovation opportunities
- Increase the probability of commercial success

No comparable dataset exists that connects natural compounds to human health outcomes with this level of scale, structure, and validation.

Brightseed's innovation platform addresses a structural challenge across life sciences: innovation is fragmented, slow, and high-risk. By integrating proprietary data with AI, Brightseed enables a continuous innovation model that shifts risk earlier, accelerates decision-making, and improves success rates.

About Brightseed

Backed by over 40 industry leading partners, Brightseed® exists to empower health science teams with the foundation needed to advance science, develop better products, and ultimately transform health through AI. The Bay Area company brings together biology, the world's largest proprietary dataset, and over a decade of computational science expertise with a shared commitment to illuminating biology and translating scientific insight into real-world impact. Brightseed's work is guided by the belief that better health outcomes require clarity, precision, and an integrated approach to innovation.

For more information, visit www.brightseedbio.com.

Media Contact

Nicki Briggs, MS, RDN
Communications, Brightseed
press@brightseedbio.com