Sikka Insights Launches Galaxy: De-Identified Real-World Healthcare Data, and a Calibrated Environment to Train Agents

Galaxy combines secure, de-identified healthcare data with calibrated training environments to accelerate safe AI agent development.

SAN JOSE, CA, UNITED STATES, September 14, 2026 /EINPresswire.com/ -- Sikka Insights, Inc. today launched Galaxy, offering de-identified retail healthcare data to AI laboratories via two products: Galaxy Corpora, for direct data licensing, and Galaxy Environments, a hosted platform where AI agents interact with practice management interfaces with outcomes calibrated on two decades of observed healthcare data.

The underlying corpus has been accumulating inside working medical practices for over 20 years and has not previously been offered to the AI research market. It comprises 1.4 billion classified files across approximately 1.5 petabytes, including 887 million de-identified dental radiographs, 192 million patient records across 46,000 dental and orthodontic practices, 65 million operational and clinical documents, and 43 million animal patient records across 2,100 veterinary practices with 1.3 billion procedure records.

Galaxy Corpora: the record, delivered

Corpora are delivered into customer infrastructure, priced by volume across five modules:
- Galaxy Imaging: 887 million dental radiographs in standard DICOM and native sensor formats, delivered alone, linked to the coded procedure that followed, or linked to the full patient record.
- Galaxy Records: 192 million patient records across six linked domains covering procedures, appointments, adjudication, and demographics.
- Galaxy Documents: 65 million practice documents including intake forms, care plans, prior authorizations, and billing logs.
- Galaxy Veterinary: 43 million animal patient records and 1.3 billion procedure records.
- Galaxy Encounters: Complete encounter packages pairing clinical documentation with filed codes and actual billing outcomes.

Galaxy Environments: the record as a system that responds

Galaxy Environments allows AI agents to be trained and evaluated against dental and veterinary

practice operations. When an agent acts inside the environment, responses and claim adjudications are sampled from 20 years of empirical, observed network data rather than artificial dynamics.

Each response includes cohort sizing and sampling distributions. Environments feature four fidelity levels—from historical encounter replays to generative and adversarial cohorts—supported by published calibration reports comparing model behavior against held-out slices.

Every response is returned together with the distribution it was sampled from and the size of the matched cohort, and each dynamics component publishes a calibration report comparing models against observed outcomes on a held-out slice. Environments operate at one of four fidelity levels, from replay of real historical encounters with known outcomes, through calibrated and generative modes, to adversarial cohorts weighted toward rare codes and denied claims.

Because no licensed record leaves Sikka Insights infrastructure under any environment configuration, Galaxy Environments is available to organizations whose compliance functions cannot approve receipt of de-identified health records at all. The company expects that to open a segment of the market that data licensing alone cannot reach.

Sikka Insights is not only a supplier to teams building healthcare agents. The Sikka platform's Self-Driving Practice, an agentic orchestration layer that runs the business operations of a dental practice, reached 2 million dollars in committed annual recurring revenue within three and a half months of launch, among the fastest ramps the company has seen in health technology AI. The operational experience behind that product informs how Galaxy Environments models what happens when an agent acts inside a working practice.

Why the underlying corpus cannot be reproduced

Real-world clinical and operational records reside inside practice software with no public web presence. They are source-coded during billing, providing an irreplaceable 20-year historical dataset that cannot be acquired through web crawling or public repositories like TCIA (~30M images) or MedAlign (275 patients).

Three kinds of buyer, one corpus

- Frontier Laboratories: Pre-training, post-training, and evaluation at corpus scale under master agreements.
- Emerging Laboratories: Tailored cohorts starting at 10,000 units, with subscription access to environments without data transfer.
- Data Partners: Wholesale, sublicense, or revenue-share structures for marketplaces, research networks, and evaluation providers.

Market activity

Sikka Insights is actively engaging AI laboratories, cloud providers, and data marketplaces across both product lines.

"The industry has spent three years running out of internet. What it has not run out of is reality," said Vijay Sikka, Founder and Chief Executive Officer of Sikka Insights, Inc. "For 20 years we have been the layer inside the software that dentists and veterinarians actually run their practices on. Every image, every procedure, every claim and every outcome was written to the same record by the same system, and every one of them is de-identified before it leaves our environment. Nobody assembled that afterwards, which is precisely why nobody else can assemble it now. Some laboratories want the record. Others want to train against it without ever holding it. We are opening it both ways."

Governance

All human records are de-identified in accordance with HIPAA (45 CFR 164.514(b) Safe Harbor) before any transfer, and no identifiable information is delivered under any Galaxy agreement. The underlying data is first-party and consent-based, collected under agreement with the practices that generate it, with consent captured per data category. Galaxy is licensed rather than sold, on time-boxed terms with a defined field of use. Under Galaxy Environments, no licensed record is transferred at all; customers receive responses, trajectories and results.

Availability

Both product lines are available now. Corpora and environments are independent agreements executing as order forms under a common master agreement, so a customer may hold either or both. Full details of both product lines are at galaxy.sikka.ai. Laboratories and data partners can request scoped cohorts, environment access, samples under mutual non-disclosure agreement and firm quotations at galaxy@sikka.ai.

About Sikka Insights, Inc.

Sikka Insights, Inc. business founded in 2019 is the de-identified data business serving the Sikka platform. Sikka has operated as the data and AI infrastructure layer for retail healthcare, integrated at read level with hundreds of practice management systems spanning 46,000 dental and orthodontic practices and 2,100 veterinary practices across the United States, Canada and Australia. The business is in its third consecutive year of profitability and has been EBITDA positive throughout. More information is at galaxy.sikka.ai.

Jimmy Tran
Sikka.ai
+1 800-947-4552
email us here

Visit us on social media:
LinkedIn
Facebook
X

This press release can be viewed online at: https://www.einpresswire.com/article/936014393

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.