



GENERATIVE INFORMATION ENGINE

GENIE

By Ayesha Imran



PROBLEMS

Complexity Overload

Existing tools for dataset creation are often overly complex, requiring extensive technical know-how and time to navigate.

High Costs

Many dataset creation solutions come with hefty price tags, making them inaccessible for quick and cost-effective model fine-tuning.

Lack of Rapid Solutions

Developers face a gap in tools that offer fast, simplified synthetic dataset generation for immediate fine-tuning and testing needs.



SOLUTIONS



Streamlined Interface

GENIE offers a clean, intuitive interface where users can effortlessly define dataset fields and generate data in seconds, eliminating complexity.

Cost-Effective

Providing a low-cost, accessible solution with a generous free tier, GENIE allows developers to create custom datasets without the burden of expensive software.

Rapid Data Generation

GENIE accelerates the dataset creation process, enabling quick fine-tuning and testing by generating synthetic data tailored to specific needs in just a few clicks.


DEMO



A photograph showing a human hand reaching out towards a robotic hand. The human hand is in the foreground, slightly out of focus, while the robotic hand is in the background, more in focus. The scene is lit with blue and purple light, creating a futuristic atmosphere. The robotic hand has visible joints and a metallic texture.

TECHNICAL ARCHITECTURE

GENIE's core tech is IBM Watson's Granite 13B Model. IBM Watson's built-in AI guardrails make ensuring reliable and robust response generation easy. The ease of directly using various large language models' APIs, including those developed by IBM Watson as well as other open source language models, makes developing your own apps easy, hassle-free, and lightning fast.

A stylized, low-poly illustration of a robotic hand. The hand is rendered in shades of blue and purple, with visible joints and segments. It is positioned in the bottom right corner of the slide, with the palm facing upwards and fingers slightly spread.

TARGET MARKET

01

AI Developers & Researchers

Professionals focused on fine-tuning and testing large language models, seeking efficient tools for custom dataset creation.

02

Startups & Small Businesses

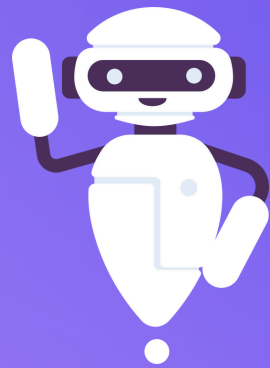
Companies looking for affordable, quick solutions to generate domain-specific datasets without investing in complex software.

03

Educational Institutions

Academics and students needing accessible tools to create datasets for research projects, experiments, and learning purposes..

PRICING PLAN



FREE TIER

\$0 per month

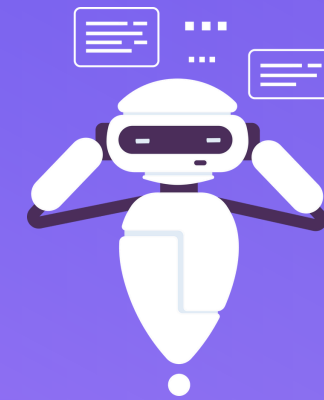
Perfect for small projects, offering up to 30 rows and 6 fields—ideal for basic dataset creation without any cost.



PRO TIER

\$15 per month

Unlock advanced capabilities with higher row limits and more fields, tailored for extensive dataset needs at an affordable price.



ENTERPRISE TIER

\$30 per month

Designed for large-scale operations, providing unlimited data generation with full customization options to meet complex requirements.

BUSINESS MODEL

FREEMIUM MODEL

Offer a free tier with limited features to attract users and demonstrate value, with paid tiers unlocking advanced features.

SUBSCRIPTION-BASED

Generate recurring revenue through subscription plans that scale with usage—targeting both individual developers and enterprise teams needing more extensive datasets.

API MONETIZATION

Provide API access for seamless integration into third-party tools, offering additional revenue streams through API usage and licensing fees.



MARKET SIZE



RAPID GROWTH

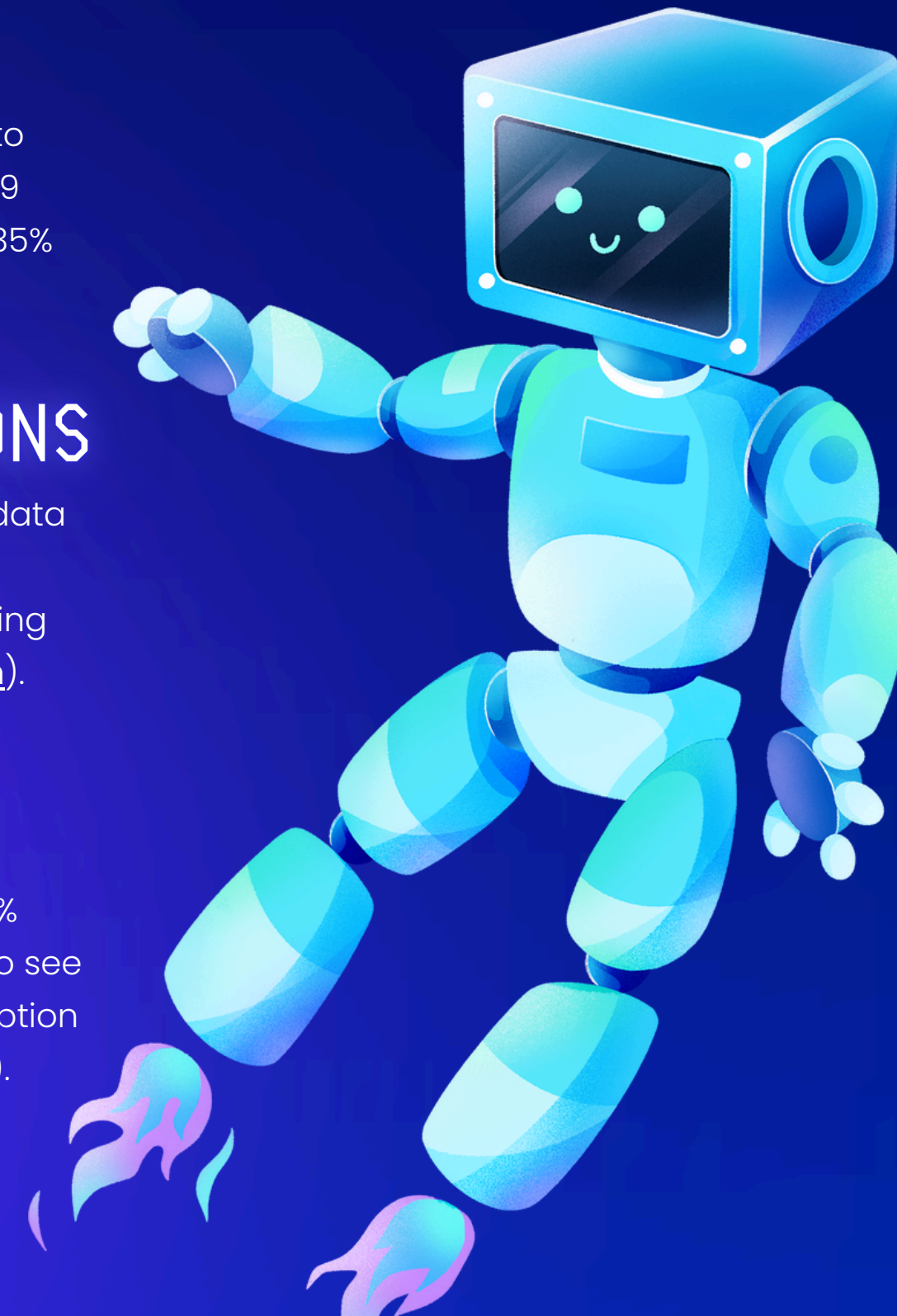
The synthetic data generation market is projected to grow from approximately \$218 million in 2023 to \$1.79 billion by 2030, reflecting a robust CAGR of around 35% ([Grand View Research](#)).

EXPANDING APPLICATIONS

With the rise of AI and machine learning, synthetic data is increasingly vital across industries, particularly in healthcare, finance, and autonomous vehicles, driving significant market demand ([Allied Market Research](#)).

GLOBAL REACH

North America currently leads the market with a 35% share, while regions like Asia-Pacific are expected to see the highest growth rates due to the increasing adoption of advanced technologies ([Allied Market Research](#)).



SAM & TAM

Total Addressable Market (TAM)

The global synthetic data generation and AI dataset creation market is expected to reach \$3.5 billion by 2031, representing the full potential revenue opportunity if GENIE were to capture 100% market share across all sectors ([Allied Market Research](#)).

Serviceable Available Market (SAM)

GENIE specifically targets AI developers, researchers, and startups focused on fine-tuning models. This segment could represent a significant portion of the SAM, estimated at \$1 billion by 2028, within the broader AI tools market.



TIMING

01

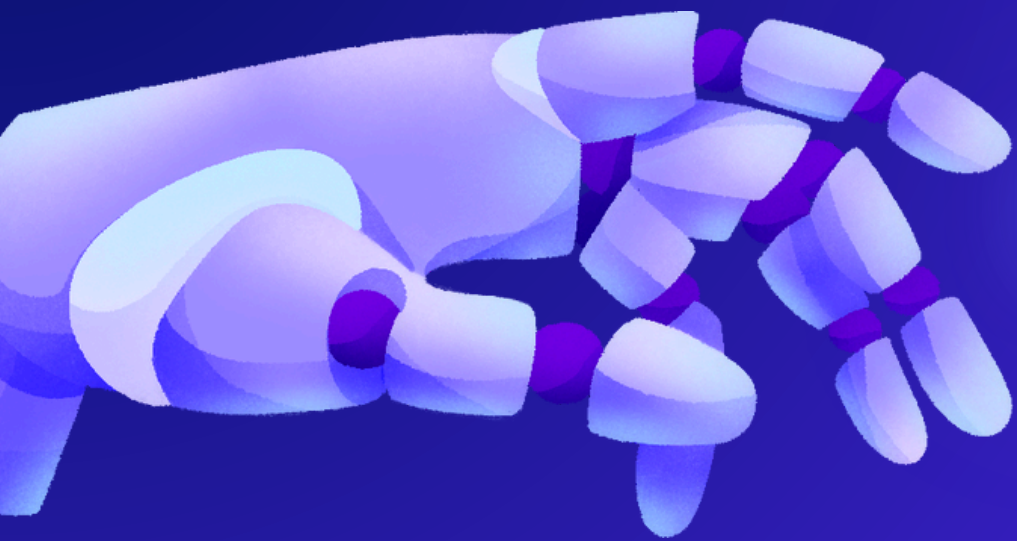
- **Surging Demand for LLM Fine-Tuning:** As businesses rapidly adopt AI, the need for custom datasets to fine-tune models has never been greater, positioning GENIE perfectly in this high-growth market.
-

02

- **Complexity Fatigue:** Developers are increasingly seeking simplified tools to manage the complexity of AI workflows, making GENIE's streamlined approach highly relevant.
-

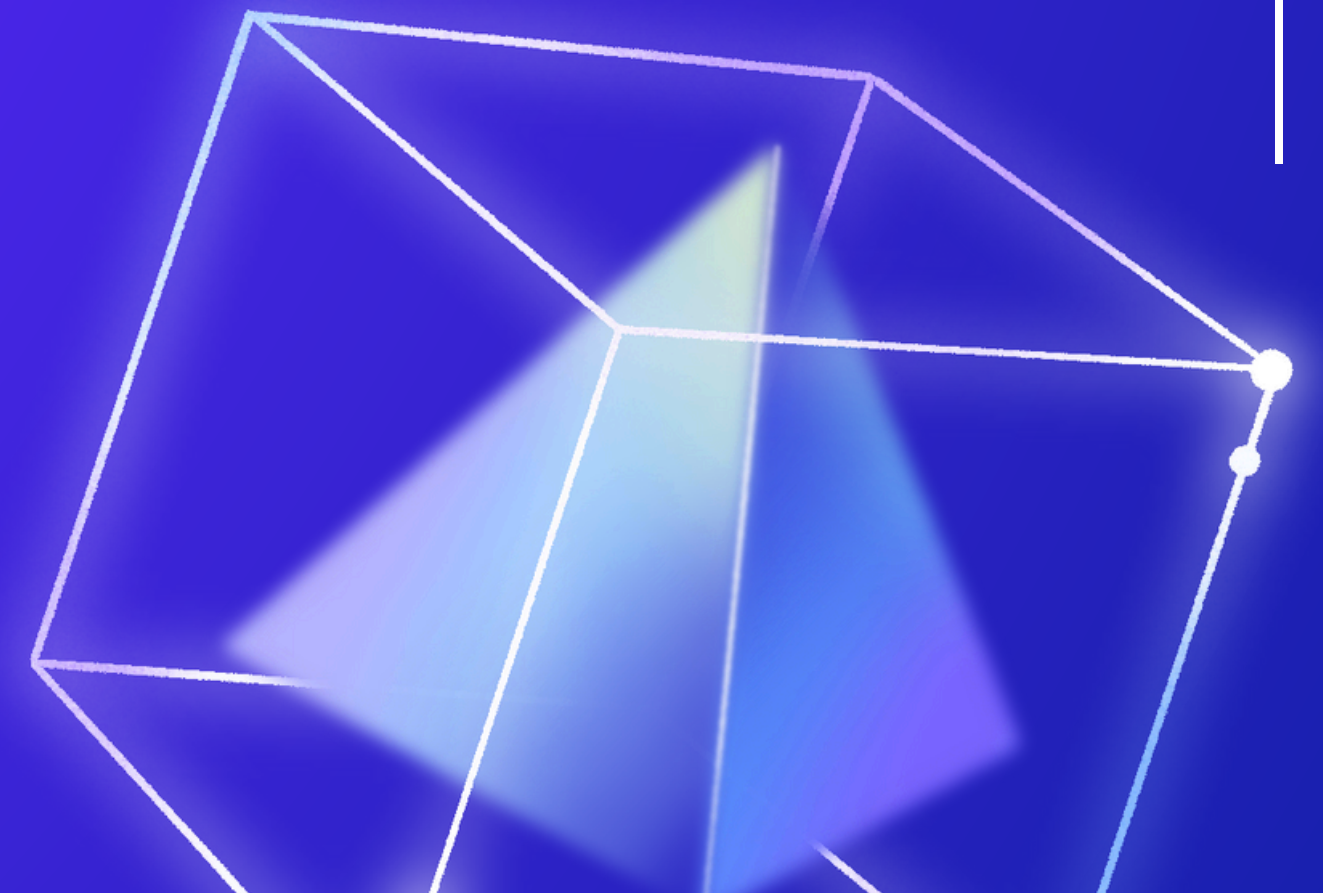
03

- **Rise of Synthetic Data:** With synthetic data generation gaining traction across industries, GENIE is launching at a time when the market is primed for easy-to-use solutions that meet this growing need.



COMPETITORS

- MOSTLY AI
- Hazy
- YData
- Statice.ai
- Edgecase.ai



COMPETITOR APPROACH



01

Complexity and Learning Curve

Competitors like YData and Hazy offer advanced synthetic data generation but come with steep learning curves, requiring significant technical expertise to configure and optimize.

02

Cost and Resource Intensive

Platforms such as Edgecase.ai and MOSTLY AI, while powerful, demand considerable computational resources and are often costly, limiting accessibility for smaller teams or individual developers.

03

Built-in Guardrails

IBM Watson's language models, integrated into GENIE, include robust guardrails that filter out harmful content, ensuring that all generated synthetic data is safe and compliant with ethical standards.

FUTURE EXPANSION



Increased Personalisation

Enable chatting generated dataset

Dataset Editing

Allow users to directly modify and refine generated datasets.

Dataset Extension

Users can upload existing datasets for further expansion by the app.

Multilingual Support

Introduce support for multiple languages in dataset generation.

FUTURE EXPANSION



Multimodal Data

Expand capabilities to generate synthetic images and audio, alongside text.

Increased Capacity

Scale up data generation beyond the current 100-row limit.

Enhanced Guardrails

Strengthen guardrails for more reliable and secure data generation.

Domain-Specific Accuracy

Improve responses in sensitive fields like healthcare, law, and finance.

FUTURE EXPANSION



Customizable Templates

Offer pre-built dataset templates for various industries.

API Integration

Provide API access for seamless integration into other platforms and tools.

Real-time Data Validation

Implement real-time validation tools to ensure data quality as it's generated.

THANK YOU!

