Snorkel

# Guide to Data Labeling & Annotation

## for Enterprise AI Teams

Accelerate and scale the delivery of high-quality AI-ready data.

DECEMBER 2024

# Table of Contents

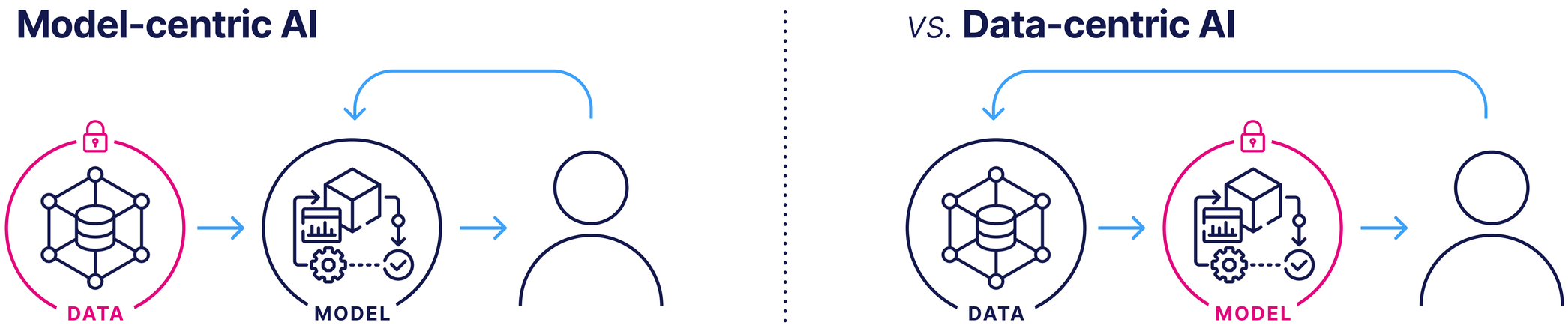
# The New Age of
# Data Labeling & Annotation

In the age of large language models (LLMs) and generative AI (GenAI), data labeling—and, increasingly, data annotation—has become a critical, complex component of enterprise AI.

These technologies require vast amounts of high-quality, context-specific labeled and annotated data to perform effectively, whether you aim to build customer service automation or a contract analysis system. Traditional methods, such as using "naturally" labeled business data, now struggle to keep up with the sophistication and scale needed to build and deliver production GenAI applications. Businesses have attempted to fill these gaps by asking employees to label data via spreadsheets, email, and Slack messages, but this has proven to be clumsy, unreliable, unauditable, and expensive.

**GenAI systems display an aggressive hunger for labeled data, which supports everything from baseline system evaluation to fine-tuning LLMs to optimizing data tagging and retrieval in vector databases for retrieval-augmented generation (RAG) systems.**

To address the elevated demand for labeled and annotated data, enterprises are turning to programmatic labeling. In this approach, data scientists and subject matter experts (SMEs) collaborate to build domain-specific rules and heuristics that apply labels and annotations at scale with minimal manual intervention. This approach accelerates data preparation and enhances consistency, making it possible to generate the high-quality labeled and annotated data that enterprise AI requires.




**This guide explores the unique challenges in modern data labeling and annotation and provides scalable solutions, programmatic labeling best practices, and real-world case studies. It offers a roadmap for enterprises to navigate the complexities of data labeling and annotation as they begin developing GenAI applications.**

# Data Labeling vs. Data Annotation

People often use the terms data labeling and data annotation interchangeably, but they serve different functions. For the purposes of this book, labeling categorizes data at a document level. Annotation functions as a specific subset of labeling that marks detailed features or entities within the document. Different applications may call for labeling, annotation, or a combination of both.

## Data Labeling

*Labeling assigns a general category, class, or tag to an entire record, such as an image, document, or text block.*

## Data Annotation

*Annotation applies labels at a granular level by highlighting and labeling specific regions, entities, or components within a document.*

## Examples of Labeling vs. Annotation

To understand the difference between data labeling and data annotation, think of pictures that may contain dogs. In a labeling project, a labeler who receives an image of two dogs would apply the category "dog." They would apply the same label no matter how many dogs appeared.

An annotation project would treat this image differently. Annotators would draw a bounding box around each dog in the image and attach a label to each box. This demands more effort than document-level labeling but yields a richer dataset appropriate to more precise AI applications.

The same principle applies to text documents and PDFs. A labeling project would categorize an entire document according to a chosen schema. An annotation project would apply labels at a word or phrase level—perhaps identifying the names of companies referenced in a contract.

## Nuances and Overlap Between Labeling and Annotation

The line between labeling and annotation can sometimes blur. If an annotator highlights a document section and applies the label "introduction," that's annotation. A RAG pipeline might take the same document, break it into smaller chunks, and apply labels to each. This could result in an identical passage receiving the same label.

Despite the identical outcome, the latter case falls into the labeling category. The pipeline treats each chunk as its own document, while the annotation project treats the section as a piece of a larger whole.

# Data Labeling in the Age of LLMs and Generative AI

GenAI systems typically contain multiple modular components (including several AI models) that collectively filter requests, retrieve relevant information, deliver responses, and tailor outputs to specific contexts. Data labeling plays a central role in this modular architecture.

Effective data labeling supports fine-tuning, the curation and annotation of documents in vector databases, and even the embedding models that power context retrieval.

## Data Labeling for RAG Systems

Traditional RAG pipelines include an orchestrator/agent, an embedding model, a vector database, and an LLM. Curating and labeling data for inclusion in vector databases, along with tagging for better filtering, is critical to ensuring that the generative model retrieves and uses appropriate context.

### 01 | Enhanced Filtering and Retrieval:

Advanced RAG systems may employ custom, ML-powered data-enrichment APIs to apply predicted labels to records within the database. These tools might label records as containing dates or definitions. This enables more granular filtering and enhanced prioritization at retrieval time, improving the quality of generated responses.

### 02 | Custom Embedding Models:

Embedding models play a crucial role in generative AI systems. They provide a means to convert “records” into vector representations, also known as embeddings. These form the foundation of similarity search and contextual matching. Embedding models fine-tuned on labeled data for particular domains—like finance, law, or healthcare—yield vectors that capture industry-specific nuances and relationships, resulting in more accurate retrieval.

**Data labeling techniques play a critical role in optimizing how vector databases index and retrieve relevant context. This, in turn, makes LLM responses more accurate and context-sensitive.**

## Curating and Labeling Data to Fine-Tune LLMs

Fine-tuning an LLM to align with enterprise requirements and corporate policy involves curating high-quality, domain-specific prompt/response pairs. Regardless of whether data scientists use reinforcement learning with human feedback (RLHF), direct preference optimization (DPO), or any other fine-tuning approach, these pairs demonstrate to the model what accurate and appropriate responses look like for the organization's specific needs and standards. In some cases, data scientists collect multiple responses to a prompt and have experts rank them.

Fine-tuning with carefully curated prompts and responses can help mitigate issues like hallucinations and irrelevant content generation by giving the model clear, contextually accurate examples to emulate. This process also ensures that the model adheres to the tone, formality, and subject matter expertise required for enterprise applications.

## Building Modular GenAI Systems with High-Quality Labeled Data

Labeling is essential for optimizing the modular components of GenAI systems. From LLM fine-tuning to embedding model training, high-quality labeled data provides the structure that these models need to perform consistently and accurately.

# Traditional Labeling Methods and Their Limitations

**Enterprises today face a new era of AI demands, one in which traditional data labeling methods struggle to keep up**

## “Natural” Labeled Data

For years, enterprises have relied on data labeled as part of natural business processes—such as transactional records in finance or churn indicators in subscription services. This naturally labeled data has driven significant value and supported numerous AI applications, but enterprises are close to exhausting its value.

Enterprises now seek to push beyond the confines of their existing datasets to power advanced GenAI and RAG applications, both of which demand higher quality, more diverse, and more precisely labeled data than traditional methods can provide.

## Manual Labeling by Subject Matter Experts (SMEs)

Manual labeling by SMEs remains one of the most accurate approaches for creating high-quality labeled data —especially for specialized, domain-specific applications.

In industries like finance, healthcare, and legal services, where labeling might involve interpreting complex legal clauses, assessing medical records, or evaluating financial disclosures, SMEs bring essential expertise that ensures accuracy and relevance. **However, this precision comes with challenges:**

### Cost and Resource Constraints:

SMEs are often highly compensated professionals, which makes extensive manual labeling projects cost-prohibitive. In addition, SMEs often resist data labeling and annotation tasks, which they may find tedious.

### Time Constraints:

Manual labeling by SMEs creates long turnaround times. Given their existing workloads and the time it takes to label complex data accurately, projects relying on SME-based labeling often experience significant delays.

### Privacy and Compliance Constraints:

Data privacy laws and security protocols can restrict access to sensitive data, limiting which SMEs are allowed to work on the project. Ensuring compliance with these regulations while still creating labeled data can impose substantial logistical and legal challenges, particularly as data privacy laws continue to evolve.

## Going Beyond Traditional Labeling and Annotation

Traditional labeling methods—natural labels and manual SME annotation—have been essential in establishing early AI models and use cases. However, as enterprises pursue more advanced AI applications like generative AI and RAG, these approaches face practical limitations around cost, scalability, accuracy, and compliance.

To meet the new data requirements of these applications, enterprises must consider more scalable, flexible solutions, such as programmatic labeling, which we will explore in the following section.

# Scalable Data Labeling Solutions for Enterprises

**To meet the evolving data needs of advanced AI applications, enterprises are adopting scalable labeling solutions that offer speed and flexibility beyond traditional approaches**

**Here we explore three prominent solutions:**

## 1 Crowd Labeling Firms

Crowd labeling firms provide enterprises with a managed, scalable way to source labels by coordinating large groups of remote annotators.

Crowd labeling firms are highly scalable and cost-effective for basic labeling tasks.

Limited domain expertise makes crowd labeling less suitable for nuanced data. Each labeler will also bring with them their own bias, which will infuse their labels. Privacy regulations often restrict enterprises from sharing data externally, limiting its use in industries with strict compliance requirements. Crowd labeling also results in rigid datasets; if the schema changes, the labeling project often must start again from zero, which could double its cost.

## 2 LLM-Assisted Labeling

Automatic labeling with large language models leverages the embedded knowledge in frontier foundation models to generate initial labels. This offers an efficient way to handle unstructured data sources like reports and customer feedback.

LLMs can process large amounts of data more rapidly and cheaply than crowd labelers. Depending on the domain and prompt template, they can even achieve relatively high baseline accuracy—though rarely production-grade.

LLM labels will reflect the biases of the LLM's training data, which could reduce accuracy, particularly in specialized contexts. Additionally, LLMs sometimes produce "hallucinations," or incorrect outputs, which need careful review. For proprietary data, these services raise similar concerns to crowd labeling; any private data labeled would leave your secure environment.

## 3 Programmatic Labeling

Programmatic labeling uses automated functions, heuristics, and models to label data at scale with minimal manual intervention. This approach allows data scientists and SMEs to encode domain-specific logic into reusable labeling functions, offering automation and accuracy.

*(See more in the "What Is Programmatic Labeling?" chapter.)*

Programmatic labeling scales efficiently, provides consistent labels, and integrates SME knowledge without repetitive manual input, making it well-suited to enterprise needs. It empowers teams to label data 10-100x faster than manual approaches and adapts flexibly if a project's schema must change.

Programmatic labeling demands highly engineered tools and platforms. Data science teams often like to build tools in-house, but an outside vendor dedicated to programmatic labeling tools is best suited for the job.

# Comparative Table of Labeling Approaches

| Feature | Internal SME Labeling | Automated LLM Labeling | Outsourced Crowd Labeling | Programmatic Labeling |
|---|---|---|---|---|
| **Cost-Efficiency** | **Low**—SMEs tend to be expensive | **High**—depending on LLM usage | **Moderate to low**—depending on the project | **High**—limited SME input is required after setup |
| **Scalability** | **Low**—limited by SME availability | **High**—rapid initial labeling | **Moderate to high**—depending on task complexity | **High**—scales with dataset size and task complexity |
| **Accuracy for Complex Data** | **High**—domain expertise ensures quality | **Low**—especially for tasks on proprietary data | **Low to moderate**—according to labeling cohort | **High**—captures SME knowledge in labeling functions |
| **Consistency** | **Moderate**—varies by SME | **Low to moderate**—depending on system setup | **Low to moderate**—requires quality control | **High**—reusable functions maintain consistency |
| **Privacy and Compliance** | **High**—strong control over data handling | **Moderate**—LLM handling requires review | **Low**—data shared with external annotators | **High**—internal or vendor-managed solution |
| **Adaptability and Flexibility** | **Low**—manual updates slow process | **High**—adjust schemas on a whim | **Low**—challenging to adjust labeling schemas | **High**—efficient function-based updates |

## Navigating Labeling Solutions for Enterprise AI

As enterprises pursue high-impact AI applications, scalable data labeling solutions like crowd labeling, LLM-based automatic labeling, and programmatic labeling each offer unique strengths. Among these, programmatic labeling stands out as a powerful option, combining the scalability of automation with the accuracy of domain expertise. This approach provides a robust balance of cost-efficiency, consistency, and adaptability, making it especially suited to enterprise demands.

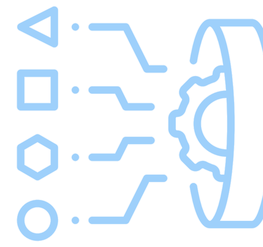
**In a later section, we'll examine the key features enterprises should prioritize in a programmatic labeling solution to fully leverage these advantages for advanced AI initiatives.**

# What Is Programmatic Labeling?

Programmatic labeling accelerates the labeling process using expert-defined rules, heuristics, and models. Users encode domain expertise into **labeling functions**—small, reusable bits of code that assign labels to data points based on specific logic. This approach allows enterprises to scale labeling efforts while maintaining accuracy, consistency, and adaptability.

## Labeling Functions: The Heart of Programmatic Labeling

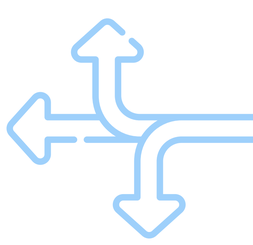
Labeling functions automate labeling by capturing specific rules or patterns, such as:



- Tagging customer reviews containing the word "excellent" as “positive.”
- Rejecting a document predicted by a model to be in the wrong language for the project.
- Sending records to an LLM-as-judge and labeling them “accept” or “reject” accordingly.

When labeling functions assign conflicting labels, a **label model** resolves these discrepancies. It evaluates each function's reliability and assigns the most likely final label to each data point. This de-noising step ensures high-quality labels even when individual functions are imperfect.

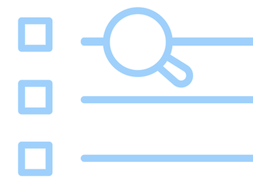
## Iterative Development for Continuous Improvement and Flexibility



Programmatic labeling is an iterative process. Data scientists analyze labeling results, refine functions, and add new ones to address gaps or errors. This feedback loop improves label quality and data coverage over time and allows datasets to evolve alongside project requirements.

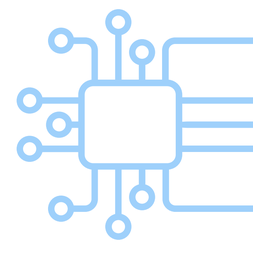
When changing needs demand a new schema, data scientists add, remove, or modify some labeling functions while maintaining those that still perform in accordance with project goals.

## Built-In Audibility



Because labeling functions are written as code, programmatic labeling is fully auditable. Users can trace labels back to their source functions, ensuring transparency and compliance—particularly important in regulated industries like healthcare and finance. Auditable workflows also simplify debugging and enable precise improvements.

## The Solution for Today’s Enterprise AI Challenges



Programmatic labeling combines scalability, accuracy, and flexibility in a streamlined workflow. By leveraging labeling functions and iterative refinement, enterprises can create high-quality datasets that power advanced AI systems while staying adaptable to evolving needs.

# Choosing a Programmatic Labeling Solution

For enterprises aiming to scale AI applications efficiently, programmatic labeling offers a unique blend of automation and domain-specific accuracy, making it well-suited for high-quality, cost-effective labeling. However, selecting a programmatic labeling solution involves balancing several factors, including scalability, SME integration, ease of use, and adaptability.

## Key Features to Look for in Programmatic Labeling Solution

01

### Customizable Labeling Functions:

An effective programmatic labeling platform allows teams to design and implement labeling functions that reflect the specific nuances of their data and domain. Look for a solution that enables flexible labeling logic.

02

### Integrated SME and Data Scientist Collaboration:

A good solution facilitates collaboration between SMEs and data scientists to develop, refine, and validate labeling functions. Some solutions even provide interfaces or workflows that allow SMEs to review and adjust outputs without extensive technical knowledge.

03

### Support for Labeling Iteration and Feedback Loops:

Advanced programmatic labeling solutions support iterative workflows, allowing teams to continuously refine labeling functions based on error analysis and model feedback. This iteration enhances data quality over time and ensures that the labeled data remains relevant to evolving business goals.

04

### Scalability and Robustness:

To support enterprise-scale labeling needs, the platform should efficiently handle large datasets without sacrificing performance. Look for solutions that offer robust performance for complex tasks, such as managing large unstructured datasets or supporting integrations with external data sources and ontologies.

05

### Security and Compliance:

Data security and compliance features are essential for enterprises in regulated industries. Choose a solution with strong access control, data anonymization, and compliance features, which can help enterprises confidently label sensitive data without risking regulatory noncompliance.

# The Snorkel / Weak Supervision Pipeline

01

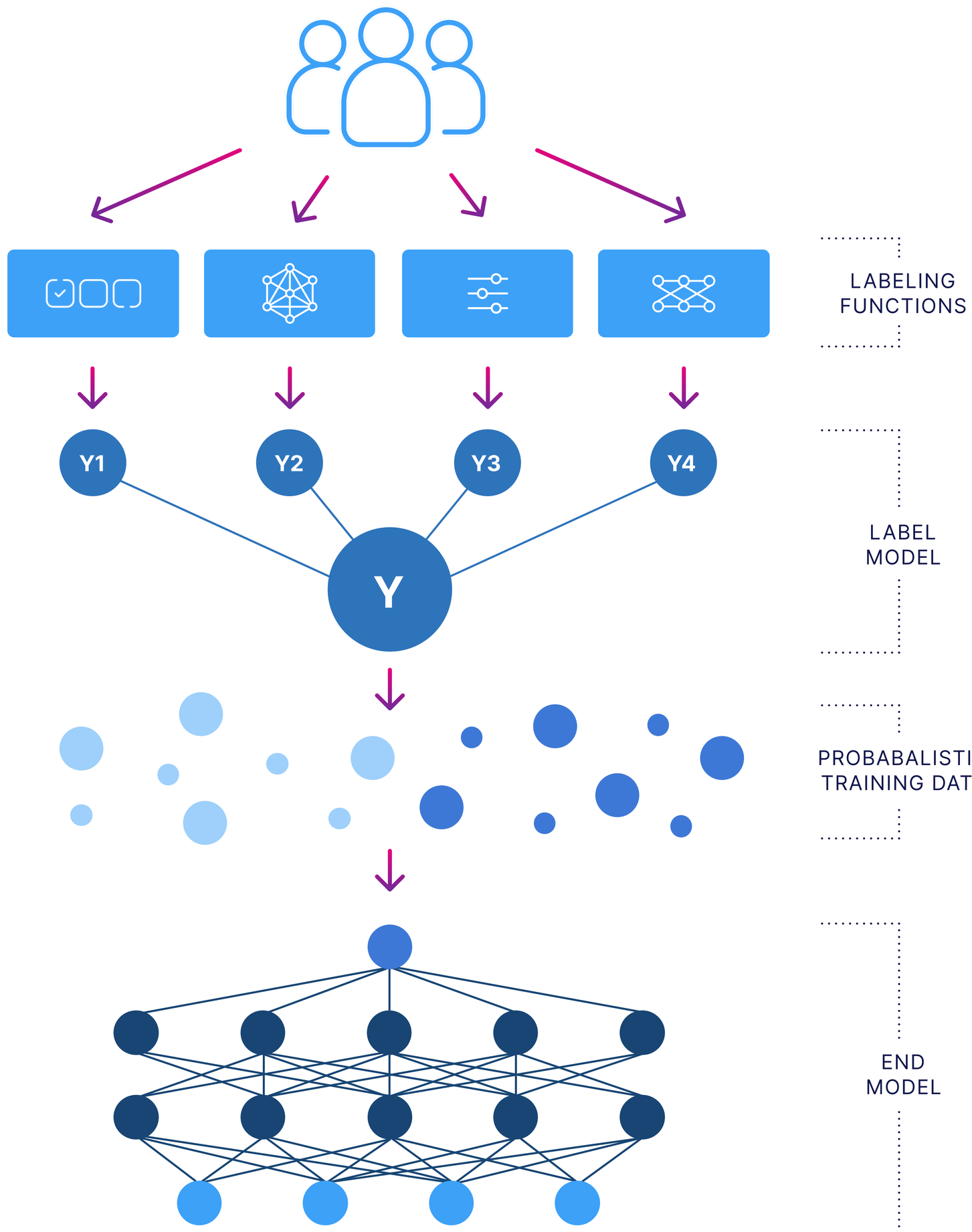
Users write **labeling functions** to create noisy labels

02

We **model and combine** these labels

03

The generated labels are used to **train a downstream model**




## Choosing the Right Solution

Enterprises can maximize their labeling efficiency by selecting a solution that matches their specific needs. While each approach has its place, programmatic labeling offers a unique combination of scale, cost-efficiency, and quality, making it ideal for enterprises pursuing advanced AI applications.

As labeling demands grow, the ability to embed SME insights into automated functions and iterate rapidly sets programmatic labeling apart as a leading solution for high-performance, enterprise-level data labeling.

# **Programmatic Labeling:**
# The Solution Your Enterprise Needs

As enterprises adopt advanced AI applications powered by LLMs and generative models, data labeling and annotation have emerged as foundational needs.

Today's systems require data that is not only labeled at scale but also aligned with the complex, domain-specific requirements of high-stakes applications. Programmatic labeling and annotation tools for enterprise AI projects offer powerful solutions by automating much of this process while embedding critical SME expertise through rules and heuristics. This approach enables enterprises to achieve the quality and specificity required without the prohibitive costs and time associated with traditional manual methods.

However, implementing and scaling programmatic labeling and annotation can be complex. It requires specialized tools that support the creation, management, and refinement of labeling functions, as well as seamless collaboration between SMEs and data scientists. For most enterprises, relying on an external provider of programmatic labeling and annotation tools is essential. These providers bring the engineering expertise, tool flexibility, and workflow support needed to ensure that data labeling and annotation operations remain efficient, adaptable, and precise.

In a landscape where speed, accuracy, and scalability are paramount, partnering with a dedicated programmatic labeling provider allows enterprises to focus on their AI-driven goals rather than the intricacies of data preparation infrastructure. As AI continues to advance, high-quality, programmatically labeled, and annotated data will remain a key enabler of competitive differentiation, powering intelligent, context-aware systems that deliver real business value.

# Case Studies

Here's a closer look at how various organizations have successfully used programmatic data labeling and **Snorkel Flow** to accelerate model development, reduce costs, and achieve higher accuracy across diverse applications.

## Global BPO Provider – *Enhancing classification and information extraction*

**RESULTS:**

- **Labeled 3 million** conversations in just **3 days**
- **Achieved 90% model accuracy** (up from 60%)
- **Accelerated model delivery by 20x**

**CHALLENGE:**

**To automate document claim reviews and sales conversation classification**, which previously required thousands of hours of SME time, the company struggled with initial model accuracy and efficiency, limiting their ability to scale.

**SOLUTION:**

By adopting Snorkel Flow, the company **replaced manual labeling with programmatic data labeling and streamlined SME-data scientist collaboration**. This allowed for the creation of high-accuracy models without the typical time costs, improving operational efficiency while enhancing customer experience.

## Healthcare Software Provider – *Streamlining medical document analysis*

**RESULTS:**

- **Saved an estimated $375K per year** in labeling costs
- **Reclaimed thousands of SME hours**
- **Improved model update times from weeks to hours**

**CHALLENGE:**

**The company aimed to classify and extract data from medical charts to enhance its risk adjustment solutions.** However, manual reviews by clinical coders were costly and slowed down their ability to deliver models at scale.

**SOLUTION:**

Using Snorkel Flow, the company **automated data labeling, created flexible label functions, and accelerated model updates**. This enabled them to adapt their training data to the needs of different healthcare clients, meeting SLAs with ease and focusing more resources on innovation.

## Large Custodial Bank – *Improving RAG contract review accuracy*

**RESULTS:**

- **Increased AI agent accuracy from 25% to 95%**
- **Accelerated data lebeling by 24x**
- **Moved from prototype to production in 6 weeks**

**CHALLENGE:**

The bank's **AI agent for contract review struggled to retrieve relevant context for LLM responses**, causing low accuracy and frequent abstentions from the AI in providing answers.

**SOLUTION:**

Snorkel AI helped **fine-tune both the LLM and embedding model using programmatic labeling and enhanced metadata extraction**, improving retrieval accuracy. This resulted in a highly effective RAG-based system that allowed SMEs to review legal contracts more efficiently and reduced compliance risks.

## **Top 3 Telecom** – *Improving customer experience with LLM labeling*

**RESULTS:**

- **Labeled 20,000** interactions in **2 days**
- **Increased model accuracy by 17 points**
- **Sped up labeling by 42x**

**CHALLENGE:**

The company's customer support **relied on an LLM-based conversational AI system that performed poorly**, leading customers to prefer speaking with representatives.

**SOLUTION:**

Using Snorkel Flow, the company **labeled high-quality agent responses to train a model that could identify actionable insights in interactions**. This fine-tuning improved the LLM's accuracy, helping deliver a better customer experience by accurately predicting response quality and actions.

## **Credit Reporting Company** – *AI copilot for customer service*

**RESULTS:**

- **Generated 300,000+** labels in **3 weeks**
- **Improved response accuracy by 6 points**
- **Achieved a 16x faster labeling speed**

**CHALLENGE:**

The company **wanted to implement a RAG-based AI copilot to help customer service representatives** (CSRs) answer inquiries accurately, but initial deployments lacked sufficient accuracy.

**SOLUTION:**

Through Snorkel Flow, the team **curated prompt-response pairs and fine-tuned both the LLM and embedding model** to improve retrieval accuracy. This resulted in a highly contextualized AI copilot that met production standards and enhanced the CSRs' efficiency in handling calls.

## **Social Media Platform** – *Scaling customer profile classification*

**RESULTS:**

- **Classified 1.5 million** customer profiles **8x faster**
- **Achieved a 11-point recall improvement**

**CHALLENGE:**

The company **needed to quickly classify many profiles** but found that existing tools could not scale, and manual labeling would have been too slow.

**SOLUTION:**

Snorkel Flow **enabled rapid iteration on training data and model updates**, allowing the team to label profiles accurately at scale. This high-recall classifier now continuously classifies new profiles, keeping up with the platform's growth.

## Snorkel Flow in Action!

These case studies illustrate how enterprises across industries are using programmatic data labeling to overcome challenges in model accuracy, scalability, and operational efficiency, driving substantial time and cost savings while improving the quality of AI outputs.

# Glossary

## Data Labeling

The process of assigning predefined categories, tags, or classifications to raw data, making it usable for training supervised machine learning models. In enterprise AI, data labeling is foundational for enabling models to learn from structured examples and generate accurate predictions.

## Data-Centric AI

An AI development philosophy that prioritizes data quality and relevance over tuning model architectures. Data-centric AI focuses on iterative data improvement (e.g., refining labels, reducing bias) to boost model performance. This approach contrasts with traditional model-centric AI and is highly applicable in enterprises where high-quality labeled data is essential for achieving robust models.

## Data Development

The comprehensive process of preparing raw data for AI, which includes cleaning, labeling, structuring, and enriching data to maximize its utility for machine learning. Data development treats data as a resource to be iteratively enhanced, often involving collaboration with SMEs and applying techniques like programmatic labeling and quality assurance to meet enterprise standards.

## Data Annotation

A specific type of data labeling where individual data points (like text, images, or audio) are marked up with metadata to enhance context and meaning. Annotations are often detailed and domain-specific, providing additional insights, such as highlighting important entities or objects in an image, that guide the model's learning process.

## Ground Truth

In machine learning, ground truth represents the most accurate data labels or classifications against which model predictions are evaluated. Establishing ground truth is critical for supervised learning as it provides the benchmark for accuracy and helps gauge a model’s performance on real-world data.

## Labeling Function

A rule or algorithm designed to apply a label to data programmatically. Labeling functions can be simple heuristics or complex scripts that encode domain-specific knowledge. They enable automated, consistent labeling across large datasets, reducing the need for manual annotation and allowing for efficient scaling in enterprise applications.

## Programmatic Labeling

A scalable approach to data labeling that uses automated labeling functions, heuristics, or machine learning models to label large datasets with minimal human intervention. In enterprise AI, programmatic labeling accelerates the creation of labeled data, enhancing speed and consistency while reducing the cost and effort of manual labeling.

## Synthetic Data

Artificially generated data that mimics real-world data attributes and patterns, often used when actual data is scarce or privacy-sensitive. Synthetic data can be created through algorithms, simulations, or generative models and is particularly valuable for training models when labeled data is limited or inaccessible due to privacy constraints.

## Weak Supervision

A machine learning technique that uses imperfect or noisy labels (e.g., from heuristics, crowd-sourcing, or programmatic labeling functions) to train models. Weak supervision allows enterprises to generate labeled datasets more quickly by combining multiple, often noisy, sources of information, enabling rapid training of high-accuracy models while balancing cost and effort.

# Key Takeaways

In today's AI-driven enterprises, effective data labeling and annotation have become crucial for scaling generative AI and LLM applications. Programmatic labeling stands out as a transformative approach that addresses the limitations of traditional methods, bringing speed, accuracy, and adaptability to labeling and annotation workflows.

## If you take away nothing else from this book, remember the following:

01

**LLM systems and GenAI applications** demand more labeled data than ever before. This data powers LLM fine-tuning and supporting models that enhance these systems' accuracy and reliability.

02

**Prioritize data labeling as a strategic asset** to ensure model accuracy and contextual relevance, particularly for high-stakes LLM and generative AI applications that depend on large, well-labeled datasets.

03

**Move beyond traditional labeling methods** to meet the scale, complexity, and compliance needs of enterprise AI. Manual and crowd-based approaches often fall short in high-demand settings.

04

**Leverage programmatic labeling for efficiency and accuracy** by combining SME expertise with automation. This approach allows enterprises to generate consistent, high-quality labels at scale, saving time and reducing costs.

05

**Partner with a specialized programmatic labeling provider** to implement robust, scalable labeling tools that offer essential infrastructure, collaboration features, and ongoing support.

**With these key strategies, enterprises can optimize data labeling to power advanced AI applications, delivering more consistent, compliant, and impactful results across industries.**

Snorkel

Ready to optimize
your data labeling?

**Get started today!**

Learn more at

**snorkel.ai**