A company based in New York City is looking for a Data Engineer on a full-time basis with the opportunity to work remotely. The company is tackling one of AI’s most foundational infrastructure challenges: secure access to high-quality training data. The company’s platform enables privacy-first and efficient data exchange across the healthcare and media sectors. With a mission-driven focus, strong early traction, and deep technical leadership, we’re building core infrastructure for the AI era.

**Why Join**

• Solve a generational AI infrastructure bottleneck from the ground up

• Work hands-on with unstructured data: PDFs, video, imaging

• High autonomy and end-to-end ownership

• Build alongside a world-class, lean engineering team

• Be part of a company shaping how AI is trained at scale

**Tech Stack**

• **Languages:** Python (primary), Java

• **Tools:** Airflow, Prefect, Dagster, DBT, Snowflake, Databricks

• **Cloud:** AWS

• **Focus Areas:** Data orchestration, compliance, scalability, lineage

**Key Responsibilities**

• Design and implement scalable data orchestration and lineage systems

• Build validation, transformation, and compliance workflows

• Optimize large-scale data storage for cost and performance

• Own projects end-to-end, from architecture to production

• Contribute as a generalist across the data platform and engineering stack

**Must-Have Qualifications**

• 6+ years of experience in data engineering (flexible for exceptional candidates with startup backgrounds)

• Exceptional Python skills

• Experience with orchestration tools: Airflow, Prefect, or Dagster

• Proficient in Snowflake or Databricks

• Strong AWS cloud experience

• Hands-on with DBT for data transformations, testing, and documentation

• Java proficiency

• Strong CS fundamentals and a technical undergraduate degree

• Experience in early-stage startups (sub-30 person teams; seed or Series A)

• Genuine excitement about the AI infrastructure movement

• Must not have too many short stints Sub 2 years

**Preferred Qualifications**

• Experience working with unstructured data (PDFs, videos, imaging)

• Familiarity with healthcare compliance (e.g., HIPAA)

• Experience at companies with strong engineering cultures

• Prior experience as a founding data engineer or early technical hire