

POLI – LLM TRAINING DATA



Text training data for your LLMs structured using an expert-led ontology and lexicon and an automated, patented metadata tagging process. Includes quantitative metadata tags that deliver context and grounding for your vector databases, knowledge graphs, and RAG processes. Spend less time creating feeds and cleaning, wrangling language data. Spend more time analyzing policy/market reaction functions

TECHNICAL DETAILS

- Full text of all official sector content in **.json format** with robust metadata tags to support a broad range of analytical & generative AI processes.
- **Lexicon & Ontology Metadata:** created by deep subject matter experts and former government officials, deployed using automated, patented metadata tagging technology.
- **Demographic Metadata:** Supports Knowledge Graph and ontology-based deployments.
- **Time Stamp Metadata:** Supports time series construction.
- **Volume-based Metadata:** Measures all four dimensions of signal in public policy language aggregate global activity across three factors (Action, Data Releases, Judicial Activity). Supports **MACD, RSI, and volatility-based analytics** regarding public policy activity; includes access to proprietary internal research and white papers. **Enhances operational efficiency and accuracy of your Vector Databases and RAG processes.**
- **Directional Metadata:** Measures policy directionality (not sentiment). in beta.
- **Analytical Metadata:** Includes Momentum Index measurements (Action x Posture).
- **Ticker Metadata:** Standard metadata tags for the entire Russell 3000. Can be customized.
- **Delivery:** Twice daily, automatically, via S3. Databricks & Snowflake upon request.
- **Coverage:** Leading advanced and emerging market economies in English (**2006-present**). Thematic Verticals: Monetary Policy, Trade, Energy, Climate, Stablecoin/Tokenization.
- **Extensions/Customization:** Keywords/Topics, Tickers, Input Sources. Tokenized Embeddings. Extensions that include deploying our patented process to tag and measure your institutional newsfeeds and other components create additional kinds of metadata tags and **support delta, theta, and gamma-based measurement of the size, duration, decay rate, and volatility of your informational advantage relative to the news cycle.**