Appendix

Data Summary

Our database combines daily PLTR stock features over time, and the number of times AI and war-related keywords appear in world news. The PLTR price data are daily OHLCV (open, high, low, close, volume) downloaded from Macrotrends [1] and saved locally. The keywords data comes from GDELT 2.0 GKG daily signals exported from BigQuery into a CSV [2]. The query counts daily AI-related theme hits (cnt_ai), war/conflict theme hits (cnt_war), and the day's average overall tone (avg_tone). These were merged on the calendar date after forward-filling the daily GDELT signals across non-trading days. The working merged file was saved locally to Colab and spans 2020-09-30 → 2025-10-03 on trading days (1,259 rows).

Provenance

This dataset combines two publicly accessible sources into a single, trading-day time series. First, we downloaded Palantir's daily OHLCV prices from Macrotrends and performed light cleaning (header normalization, date parsing, numeric coercion) before deriving returns. Second, we queried the GDELT 2.0 Global Knowledge Graph (BigQuery) to create daily news features by counting records whose V2Themes contained AI/ML/LLM terms (cnt_ai) and war/conflict terms (cnt_war), and by extracting the overall tone as the first numeric element of V2Tone (avg_tone). We then expanded the GDELT features to a daily calendar and forward-filled them across weekends/market holidays before left-joining to PLTR trading days, producing one consolidated file used for EDA and modeling.

License

OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER
DEALINGS IN THE SOFTWARE.

Ethical Statements

This dataset contains no personal data. Interpretation risks: model outputs could be misconstrued
as investment advice; news coverage has editorial/geographic/language biases that can affect
cnt_ai/cnt_war. We document our theme lists and use the signals for exploratory/educational
modeling only.

Data Dictionary

Field	Type	Description	Notes / Uncertainty	Example
date	datetime	Trading day (NY market calendar).	Non-trading days omitted after merge.	2024-11-15
open, high, low, close	float	Daily OHLC prices (USD).	Likely unadjusted close (splits/dividends not applied).	open=21.34, high=22.10, low=20.95, close=21.87
volume	int/float	Daily shares traded.	Units = shares.	38,245,100
return	float	Simple daily return of close.	Undefined on first row.	0.012 (+1.2%)
log_return	float	log(close).diff().	Good for additivity/stationarity.	0.0119
cnt_ai	int	Daily count of GKG rows with AI/ML/LLM themes.	Keyword coverage may evolve; regex documented in query.	57
cnt_war	int	Daily count of GKG rows with war/conflict/military themes.	Much higher baseline than AI globally (expected).	512,304
avg_tone	float	Average overall V2Tone for that day.	Centered near 0; negative values common.	-1.62

Exploratory Plots

PLTR Close Price Over Time

Daily AI-related Theme Count (GDELT GKG)