commodity\_analysis\_rules\_v1.0.md

# 1. Input File Schema

Required columns (in order):

1. Publish Date (YYYY-MM-DD, ISO 8601)  
2. Headline (plain ASCII, ≤120 characters)  
3. Extract (plain ASCII, ≤500 characters)  
4. Article Click-Through URL (valid http/https URL)  
5. Language (ISO 639-1 code, e.g., "en", "zh")

Validation formula:

VALID = AND(  
 ISDATE(Publish\_Date),  
 LEN(Headline)>0,  
 LEN(Headline)<=120,  
 LEN(Extract)>0,  
 LEN(Extract)<=500,  
 ISTEXT(Article\_URL),  
 REGEXMATCH(Article\_URL,"^https?://"),  
 REGEXMATCH(Language,"^[a-z]{2}$")  
)  
IF NOT VALID ⇒ Critical\_Error

# 2. Approved Website/Source List

Use only reputable, independent sources for fact-checking and validation. At least two independent sources must confirm any key claim.  
Approved sources include:  
- Reuters  
- Bloomberg  
- Financial Times  
- Wall Street Journal  
- CNBC  
- S&P Global  
- Platts  
- Argus Media  
- Nasdaq  
- London Metal Exchange (LME)  
- U.S. Geological Survey (USGS)  
- International Energy Agency (IEA)  
- Official government/industry sites (e.g., US Department of Energy, World Bank)  
- Company press releases (must be corroborated by a second independent source)  
If a source is not on this list, escalate for review.

# 3. Event Categories & Subcategories

Event Category (choose one):  
1. Price Fluctuation  
2. Supply and Demand Balance  
3. Competitive Advantage  
4. Regulatory Change  
5. Business News  
6. Irrelevant  
7. Duplicate

Event Sub-categories (choose one per category):  
- Price Fluctuation: Price Increase, Price Decrease, Forecast - Higher Price, Forecast - Lower Price, Increase < 10%, Decrease < 10%, Other  
- Supply and Demand Balance: Supply Increase, Supply Decrease/Shortage, Demand Increase, Demand Decrease, Forecast - Higher Demand, Forecast - Lower Demand, Forecast - Higher Supply, Forecast - Lower Supply, Other  
- Competitive Advantage: Auction, New Capacity, Surplus Supplies, Other  
- Regulatory Change: New Regulation, Regulation Update, Quota Increase, Quota Decrease, Other  
- Business News: Business Sale, Merger & Acquisition, Bankruptcy, Profit Warning, Shutdown/Production Cut, Raw Material Disruption, Cargo Disruption, Force Majeure, Other  
- Irrelevant: Not related to price, supply, demand, or competitive/market factors  
- Duplicate: Previously covered event

# 4. Moving Factors

Select one or more from this closed list:  
- Higher Central Bank Investment  
- Lower Central Bank Investment  
- Higher Dollar Value  
- Lower Dollar Value  
- Higher Interest Rate  
- Lower Interest Rate  
- Higher GDP  
- Lower GDP  
- Geopolitical Unrest  
- Geopolitical Stability  
- Market Uncertainty  
- Market Fear  
- Market Confidence  
- Manufacturing Disruption  
- Manufacturing Easing  
- Shipping / Cargo Disruption  
- Shipping / Cargo Disruption Easing  
- Raw Material Disruption  
- Raw Material Easing  
- Labor Disruptions  
- Labor Disruptions Easing  
- Higher Consumption  
- Lower Consumption  
- Seasonal  
- Stockpiling  
- Law & Regulation  
- Higher Inflation  
- Lower Inflation  
- Balance Competition  
- Anti-dumping  
- Higher Tax, Fee, or Fine  
- Lower Tax, Fee, or Fine  
- Higher Operational Cost  
- Lower Operational Cost  
- Financial Instability  
- Financial Stability  
- Higher Sourcing Options  
- Lower Sourcing Options  
- Stricter Environmental Laws  
- Easing Environmental Laws  
No additions allowed unless formally approved.

# 5. Currency & Unit Conversion

Currency:  
USD\_Value = Foreign\_Value × LiveRate  
Where LiveRate is the latest exchange rate from a reputable financial source (e.g., https://api.exchangerate.host/latest?base=[FX]).  
Append “(converted)” to the value if conversion is performed.  
Example: "A$8M" at 0.65 rate → "USD 5.2M (converted)"  
  
Temperature:  
F = (C × 9/5) + 32  
Example: 25°C → 77°F  
  
Distance:  
mi = km × 0.621371  
Example: 100 km → 62.1371 mi  
  
Mass:  
short\_tons = tonnes × 1.10231  
Example: 10 tonnes → 11.0231 short tons  
  
If any live rate or conversion factor is unavailable, escalate for review.

# 6. Output Format (Schema)

All fields must be present, in this order:  
1. Date  
2. Headline  
3. Commodity  
4. CASid  
5. Activity  
6. Event Category  
7. Event Sub-category  
8. Moving Factor  
9. Type  
10. Predictive Nature  
11. Prediction  
12. Prediction Maturity Period (Estimated days)  
13. Potential Impact  
14. Article URL  
15. Summary  
16. Confidence Rating  
17. Confidence Reason

# 7. Field-by-Field Rules & Formulas

Date: Date = TEXT(Publish\_Date,"yyyy-mm-dd")  
Headline: ASCII only. If facility present: "[Action/Subject] [Facility/Event] in City, State, Country". No dashes or curly quotes.  
Commodity: Use NER to extract. If multiple, list all separated by “;”.  
CASid: Lookup from official CAS DB for each commodity.  
Activity: Extract main verb from Headline (e.g., “Shutdown”, “Launch”, “Tax Hike”).  
Event Category: Assign from closed list above.  
Event Sub-category: Assign from closed sub-category list for chosen category.  
Moving Factor: Select all explicit or strongly implied (cosine similarity ≥ 0.75) from list above.  
Type: If “shortage”, “cut”, “disruption” → “Risk”. If “new capacity”, “surplus”, “easing” → “Opportunity”. If no forecast → “Market News”. If not relevant → “Irrelevant”.  
Predictive Nature: “Yes” if text includes “will”, “expected”, “forecast”, “projected”, else “No”.  
Prediction: If Predictive Nature = Yes, assign: Lower Supply Higher Price, Higher Supply Lower Price, Lower Demand Lower Price, Higher Demand Higher Price.  
Prediction Maturity Period: Extract integer days from text; bucket into 10, 30, 60, 90+. If not explicit but “month” present, default to 30.  
Potential Impact: Compute: Scope\_Score = RegionWeight × AffectedVolume%. Map: 0-2 = Low, 2-4 = Medium, 4-6 = High, ≥6 = Severe. Where RegionWeight: Local=1, National=2, Regional=3, Global=4.  
Article URL: Pass-through, must be valid http/https URL.  
Summary: Template: "Commodity – key event … (Source\_Name)". Max 75 words. Start with commodity, end with source.  
Confidence Rating: High: 2+ independent sources, no ambiguity. Medium: 1 source or minor ambiguity. Low: speculative/unverified.  
Confidence Reason: 1-sentence justification referencing source count & ambiguity.

# 8. Confidence Rating Logic

High: 2+ independent sources, no ambiguity  
Medium: 1 source or minor ambiguity  
Low: Speculative or unverified

# 9. Error Handling

If any rule is violated, output:  
ERROR: Rule violation in row [row\_id] — [reason]  
Log the error and halt further processing for that row.

# 10. Example Output Row

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Headline | Commodity | CASid | Activity | Event Category | Event Sub-category | Moving Factor | Type | Predictive Nature | Prediction | Prediction Maturity Period | Potential Impact | Article URL | Summary | Confidence Rating | Confidence Reason |
| 2024-05-01 | Shutdown at XYZ Mine in Perth, WA, Australia | Nickel | 7440-02-0 | Shutdown | Business News | Shutdown/Production Cut | Labor Disruptions | Risk | Yes | Lower Supply Higher Price | 30 | High | https://reuters.com/xyz | Nickel – Shutdown at XYZ Mine in Perth, Australia impacts supply. (Reuters) | High | Multiple sources, clear. |

# 11. Versioning

This file is versioned. Always use the latest version.