KIEL Institut

Carsten Brockhaus, Julian Hinz, Irene Iodice

April 2026

PLAID
Product-Level AI-Derived Indicators Database for International Trade

Kiel Working Paper

No. 2318, ISSN 1862-1155
Kiel Institute for the World Economy

Abstract

Carsten Brockhaus, Julian Hinz, Irene Iodice

We introduce PLAID (Product-Level AI-Derived Indicators Database), an open, versioned database of product-level indicators for international trade, generated by an ensemble of frontier large language models (LLMs). The current beta covers six indicators at the HS 6-digit level across all major HS revisions since 1992: (i) the Rauch classification of goods into organized-exchange, reference-priced, and differentiated products; (ii) UN Broad Economic Categories (capital, intermediate, consumption); (iii) economic perishability on a five-class scale; (iv) a hazardous-materials flag; (v) microchip/semiconductor content; and (vi) 3TG conflict-mineral content. We validate each indicator against established external benchmarks and demonstrate its empirical relevance through a targeted gravity application. PLAID is designed to grow: future releases will add indicators for technology intensity, R&D intensity, regulatory sensitivity, and other product attributes. All data and code are publicly available.

Keywords: Product-level Indicators, Large Language Models, International Trade, Gravity Models, Open Data
JELs: F14, F17, C80, C38

Authors

Carsten Brockhaus
Kiel Institut and University Kiel
carsten.brockhaus@kielinstitut.de
www.kielinstitut.de

Julian Hinz
Kiel Institut and Bielefeld University
julian.hinz@uni-bielefeld.de
www.uni-bielefeld.de

Irene Iodice
Bielefeld University
Irene.iodice@uni-bielefeld.de
www.uni-bielefeld.de

1 Introduction

Empirical research in international trade increasingly relies on product-level panel data. Granular bilateral trade flows at the HS 6-digit level are now available for most countries and years through sources such as UN Comtrade and BACI (Gaulier and Zignago, 2010). Yet the indicators used to characterize those products have not kept pace. Classical measures of market structure (Rauch, 1999) and end-use (United Nations. Statistical Office, 1971) were constructed for earlier classification vintages and have not been systematically updated as the HS has evolved.[1]

The problem takes two distinct forms. For classical indicators, the issue is degrading coverage and cross-revision inconsistency. Rauch's (1999) classification of goods by price-formation institutions remains one of the most widely used product-level measures in trade research, despite being defined only for SITC Rev. 2 and never updated. Mapping it to HS codes via standard crosswalks introduces noise: the same physical product may receive different Rauch labels depending on the HS revision used, while in HS 2022, 69% of 6-digit codes cannot be matched to a Rauch category through any standard crosswalk.[2] The UN Broad Economic Categories have fared somewhat better — the 2016 Rev. 5 restructuring was a genuine product level revision — but coverage of HS 2017 and HS 2022 codes still relies on correspondence tables rather than fresh product-level assignments. Crucially, both classifications were originally constructed by human experts reading product descriptions and applying a fixed set of conceptual rules — a task that large language models can now perform at scale.[3]

For novel product attributes, the problem is more fundamental: no concordance exists at all. Researchers interested in product characteristics such as semiconductor content, economic perishability, cold-chain requirements, or hazardous-material status are therefore left to build ad hoc proxies, often at coarser levels of aggregation, or to omit these dimensions entirely. For example, studies of "time sensitivity" often classify only narrow subsets of products using external storage-life information for selected product groups (Djankov et al., 2006, 2010), or construct indicators from highly time-sensitive product descriptions rather than relying on a systematic HS6-wide attribute (Hummels and Schaur, 2013). For hazardous-materials exposure, researchers typically map transport-safety regulations onto trade codes manually, with coverage limited to narrow chemical and fuel chapters and no systematic treatment of the remaining product space (Versino and Cojazzi, 2012; Stewart,

[1]As of 3 April 2026, Rauch (1999) has approximately 4,377 Google Scholar citations. BEC is an official UN statistical classification first issued in 1971 and revised multiple times; the current operational BEC Rev. 5 was adopted in 2016 and is defined in terms of HS 2012.

[2]We document this heterogeneity systematically in Section 3.1.

[3]For borderline products that can serve either as intermediate inputs or as final consumption goods, the UNSD acknowledges that description-based judgment alone is insufficient and that empirical data (supply-use tables, household surveys) may be needed — a limitation shared by any text-based classification, including ours, and one that our ensemble confidence score is designed to flag.

2026).

This paper introduces PLAID (Product-Level AI-Derived Indicators Database), an open, versioned database designed to address both problems at their root. Rather than constructing indicators by hand and leaving them to age, we build a replicable pipeline in which large language models (LLMs) perform the same classification tasks that human experts once performed — but scalably, consistently, and for any HS revision. Each indicator is framed as a structured text-classification task: given an official HS product description and its hierarchical chapter context, an ensemble of frontier LLMs assigns labels via a constrained JSON schema. Aggregating predictions across multiple frontier models, we obtain scalable labels together with natural measures of uncertainty. Because the method operates directly on product descriptions, it can be applied consistently across HS revisions and extended automatically to newly introduced codes.

The current beta release (v0.1) provides full HS 6-digit coverage across all major revisions since 1992 for six indicators:

1. **Rauch classification** — whether a good is traded on an organized exchange, reference-priced, or differentiated, following Rauch (1999);
2. **Broad Economic Categories (BEC)** — we replicate the SNA end-use dimension of BEC, the capital/intermediate/consumption decomposition that is most widely used in empirical trade research end-use classification into capital, intermediate, or consumption goods, following the UN BEC framework (United Nations. Statistical Office, 1971);
3. **Perishability** — economic perishability on a five-class scale, from ultra-perishable (days) to non-perishable (decades), with a numeric half-life estimate;
4. **Hazardous materials** — whether a product is subject to dangerous-goods transport regulations (GHS, IMDG/IATA-DGR);
5. **Microchip content** — whether a product contains or embeds a semiconductor as a functional component;
6. **3TG conflict minerals** — whether a product is or contains tin, tantalum, tungsten, or gold, with identification of the specific mineral.

For each indicator we validate the LLM output against an established external benchmark and demonstrate empirical relevance through a targeted gravity application. For existing product attributes we also show that classifying directly at the HS-6 level outperforms concordance-based approaches that inherit coarser labels. PLAID is designed to grow: additional indicators are under active development and will be added in future releases. All data are publicly available at `https://plaid.julianhinz.com` and `https://trade.ifw-kiel.de/PLAID/`.

The remainder of the paper proceeds as follows. Section 2 describes the shared LLM pipeline. Section 3 presents each of the six beta indicators with validation results and gravity applications. Section 4 describes the PLAID database format, versioning, and access. Section 5 concludes.

2 Methodology

This section describes the shared pipeline used to generate all indicators.

2.1 Data Preparation

We assemble the complete roster of HS 6-digit codes for all major revisions (1992, 1996, 2002, 2007, 2012, 2017, and 2022), extracting official product descriptions from the WCO/WTO database. For each code, we also extract the chapter-level (HS 2-digit) description to provide hierarchical context. This yields approximately 5,000–5,600 product codes per revision.

2.2 Prompt Architecture

All prompts share a common four-block structure:

1. **Role and task definition** — a system-level instruction establishing the model's expertise (e.g., trade economist, hazardous materials specialist) and a precise statement of the classification task, including the allowed output categories and their definitions.
2. **Classification guidance** — decision rules, tie-breakers for borderline cases, and positive and negative examples of each category. This block replicates the kind of annotator instructions that would be given to a human coder.
3. **Few-shot examples** — between three and five labeled input-output pairs covering the full range of output categories, including at least one borderline or ambiguous case with appropriately lower confidence.
4. **Output schema** — a strict JSON schema specifying allowed keys, value types, and permitted values. Every prompt requests at minimum: a short product description, the classification label, a 1–2 sentence reasoning field grounded in the relevant regulatory or economic concepts, and a numeric confidence score in $[0, 1]$.

The input block appended to each prompt contains the HS code, its official 6-digit description (truncated to 260 characters), and the corresponding 2-digit chapter description as hierarchical context. All prompts use `temperature=0` for maximum consistency in responses.

2.3 Output Fields and Indicator-Specific Extensions

Beyond the common fields, individual indicators include extensions suited to their classification task. The Rauch prompt adds an `evidence_type` field (one of `exchange_listing`, `reference_prices`, `differentiation_markers`, `heuristic_guess`) and a `flags` field for ambiguous cases such as insufficient product detail or conflicting grade signals. The perishability prompt similarly includes `evidence_type` and `flags`, and additionally requests a continuous `half_life_days` estimate. The 3TG prompt requests a `specific_mineral` sub-field identifying which of the four regulated minerals is present. These extensions provide auditable reasoning trails that go beyond the scalar confidence score.

2.4 Multi-Model Ensemble

A single LLM reflects idiosyncratic training data and alignment choices. To reduce model-specific bias, we query an ensemble of frontier models via the OpenRouter API:[4]

Let $c_{km} \in \mathcal{C}$ denote the label assigned by model m to product k. We construct:

$$\hat{c}_k^{\text{majority}} = \arg\max_{c \in \mathcal{C}} \sum_{m=1}^{M} \mathbf{1}\{c_{km} = c\},$$

$$u_k = 1 - \max_c \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\{c_{km} = c\},$$

where u_k is an uncertainty measure (one minus the majority share). In the main analysis we use $\hat{c}_k^{\text{majority}}$ and retain u_k for robustness checks.

2.5 Parsing and Validation

All API responses are parsed as JSON and validated against the expected schema. We drop or re-query responses that contain invalid categories, fail JSON syntax checks, or violate logical constraints. Each product's raw LLM response is stored as an individual JSON file for full reproducibility.[5]

3 Indicators

This section presents each of the six beta indicators. The first two — the Rauch classification and Broad Economic Categories — reconstruct and extend *existing* product-level

[4]Current ensemble: Mistral `mistral-small-2603`, Google `gemini-2.5-flash`, Anthropic `claude-haiku-4.5`, and OpenAI `gpt-5.4-mini`. All accessed via the OpenRouter API.

[5]The beta (v0.1) prompts were staged through a CSV pipeline whose RFC-4180 quote escaping was not collapsed when the prompt file was read back in; this produced doubled quotes inside the few-shot JSON examples that the frontier models handled gracefully (99.97% JSON parse rate). Appendix B shows the intended prompt structure matching the source scripts, which will be regenerated without the encoding artifact in v0.2.

measures that were originally defined for older nomenclatures and have not been systematically updated. We validate these against the original source concordances at their native classification vintage and show that our LLM-based classifications scale consistently across HS revisions. The remaining four indicators — perishability, hazardous materials, microchip content, and 3TG conflict minerals — are *new* product attributes that have never before been classified systematically at the HS-6 level. For these, we validate through a combination of external benchmarks (where available) and chapter-level face-validity tests, and demonstrate empirical relevance through targeted gravity applications where the economic prediction is sharp. For each indicator, we show a stylized prompt excerpt; the full verbatim prompts are reproduced in Appendix B.

3.1 Rauch Classification

Definition. Following Rauch (1999), we classify each HS 6 product into one of three categories based on the institutions of price formation: *organized-exchange* (w) goods traded on major commodity exchanges with standardized contracts and publicly quoted prices; *reference-priced* (r) goods not listed on exchanges but with widely published benchmark prices in trade journals or price-reporting agencies; and *differentiated* (n) goods where neither exchange listing nor standard reference prices exist and valuation depends on brand, design, or specifications. The classification follows a strict decision order: a product is first tested for exchange listing; if that fails, for the existence of widely published reference prices; and otherwise assigned to the differentiated category. For borderline cases, the prompt applies a tie-breaker that favors the more standardized category. The model is instructed to use the most specific grade or form implied by the HS-6 description rather than the broad heading, and to flag cases where product detail is insufficient to classify with confidence.

```
You are a trade-economics specialist.
Task: Classify each item into three categories: w, r, or n.
Definitions:
w: organized-exchange: homogeneous goods with standardized
contracts and public exchange prices (e.g., CBOT, LME, NYMEX).
r: reference-priced: not exchange-listed, but widely published
benchmark prices exist in trade journals or price-reporting
agencies.
n: differentiated: no exchange listing and no widely accepted
benchmark prices; valuation depends on brand, design,
specifications.
Decision order: exchange test (w) -> reference-price test (r)
-> else (n). Prefer the more standardized category on a tie.
Product code: HS 100590
Description: "Maize (corn)"
chapter_hint: 10 -- Cereals
```

```
Return JSON: {"rauch_category": "w",
"evidence_type": "exchange_listing",
"confidence": 0.94, "flags": []}
```

The full prompt, including all few-shot examples and the output schema, is reproduced in Appendix B.1.

Validation. We validate the ensemble on the native classification vintage used by Rauch (1999): SITC Rev. 2 at the 4-digit level. A two-model pilot (`openai/gpt-5` and `anthropic/claude-3.5-sonnet`, November 2025) generates a majority-vote consensus for the 784 SITC Rev. 2 4-digit codes that overlap with the Rauch concordance. Inter-model agreement in the pilot is 88.0% and the resulting class distribution (n=415, r=323, w=46) closely tracks the original.

Table 1 compares this consensus to Rauch's conservative and liberal variants. Overall agreement is 79.1% under the conservative coding and 75.5% under the liberal coding. Performance differs sharply by category. For reference-priced (`r`) and differentiated (`n`) goods, per-category agreement is high: 87.4% and 80.3% respectively under the conservative variant. The organized-exchange category (`w`) is substantially harder, with agreement of 49.4%; of the codes the ensemble assigns to `w`, 82.6% match the Rauch benchmark. Of the 39 misclassified organized-exchange (`w`) codes, 37 (95%) are assigned to `r` and only 2 (5%) to `n`. The ensemble thus almost never makes large classification jumps: errors are concentrated at the `w/r` boundary rather than reflecting a fundamental misunderstanding of product characteristics. This pattern is consistent with ambiguity in the original concordance itself, as the conservative and liberal Rauch variants disagree on more than one third of products. The mismatches are concentrated in a set of borderline groups — especially agricultural commodities, vegetable oils, and petroleum products — where exchange trading and reference pricing plausibly coexist. The confusion matrix is shown in Figure 1.

Table 1: Rauch SITC Rev. 2 validation (2-model pilot vs. Rauch 1999)

	Conservative	**Liberal**
N matched	702	702
Overall agreement	79.1%	75.5%
Per-category agreement		
w (organized exchange)	49.4%	36.5%
r (reference priced)	87.4%	83.3%
n (differentiated)	80.3%	83.1%
Pilot inter-model agreement	88.0%	

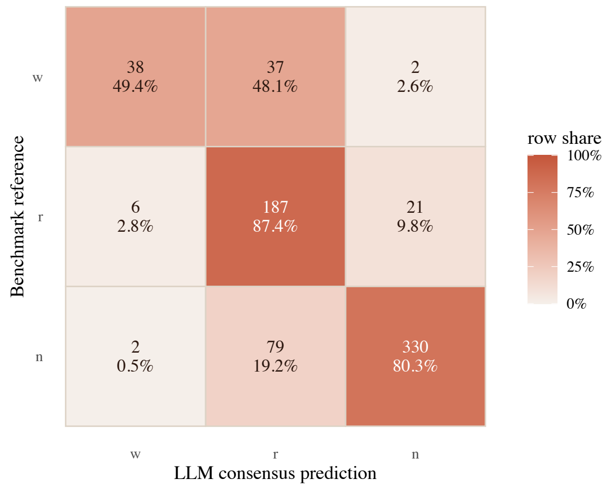

Figure 1: Rauch SITC Rev. 2 pilot: confusion matrix vs. the Rauch (1999) conservative concordance. Rows are the Rauch reference; columns are the ensemble consensus. Cells show counts and row shares. Organized-exchange (`w`) is the hardest category; errors concentrate at the `w`/`r` boundary.

Extending to HS across revisions. Because the primary validation is run on the source classification vintage, it isolates LLM accuracy from any HS concordance noise. To show that the ensemble scales across HS revisions, we map the full HS production classifications back to SITC Rev. 2 through the WITS concordances and compute per-revision agreement with the Rauch conservative benchmark. Table 2 reports agreement of 0.785 at H0 through 0.795 at H4 (the peak), with a modest decline to 0.770 at H5 and 0.776 at H6. Part of the H4→H5 gap is driven not by deteriorating LLM performance but by concordance sparsity: the number of SITC Rev. 2 codes with a complete HS mapping drops from 414 at H4 to 370 at H5, so cross-revision comparisons should be interpreted with that caveat in mind.

Table 2: Rauch HS→SITC agreement by HS revision (ensemble consensus vs. Rauch liberal)

Revision	N matched	Agreement
H0	419	0.785
H1	419	0.785
H2	419	0.783
H3	418	0.789
H4	414	0.795
H5	370	0.770
H6	371	0.776

Uncertainty as a practical screen. Across the four-model HS ensemble, agreement with the Rauch benchmark declines monotonically with ensemble disagreement. Figure 2 shows this for the HS92 codes that can be mapped to Rauch's SITC concordance: unanimous classifications (all four models agree) achieve 83.4% agreement, three-of-four agreement drops to 57.6%, and evenly split codes to 47.6%. Mean uncertainty u_k is nearly three times higher among misclassified than correctly classified products (0.15 vs. 0.05). This supports the use of u_k as a practical screening device — users can restrict to the unanimous subsample (77% of all HS-6 codes) as a robustness check.

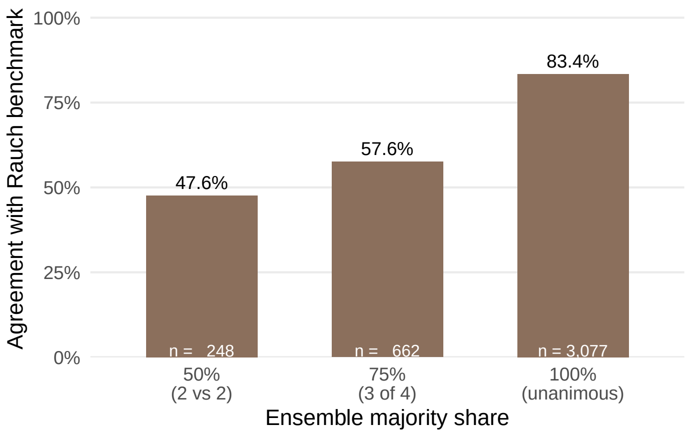

Figure 2: Ensemble agreement as a quality screen. Agreement with the Rauch (1999) conservative benchmark by ensemble majority share, for HS92 codes mapped to SITC Rev. 2 via concordance. Unanimous classifications (100% majority share, $u_k = 0$) achieve 83.4% agreement; split decisions drop to 47.6%.

Within-SITC heterogeneity. A key advantage of classifying at the HS-6 level rather than inheriting Rauch labels via concordances is that it reveals within-heading heterogeneity that concordance-based approaches mask entirely. Figure 3 illustrates this by mapping each SITC Rev. 2 code to its constituent HS-6 products and showing the distribution of Rauch categories within each SITC heading. The result is striking: a large share of SITC headings contain HS-6 codes assigned to *different* Rauch categories. For example, a single SITC heading covering "petroleum products" may combine HS-6 codes for standardized fuels (traded on exchanges, classified as `w`) with codes for specialty lubricants or blends (differentiated, classified as `n`). Under a concordance-based approach, all of these codes would receive the same Rauch label, whichever the crosswalk happened to assign. Under the LLM-based approach, each HS-6 code receives its own classification based on the product description, and the within-heading variation becomes directly measurable.

This matters for empirical work. When a gravity regression interacts distance with the Rauch category, within-heading heterogeneity introduces measurement error that attenuates the estimated differences between categories. The attenuation is larger for SITC headings that bundle exchange-traded and differentiated goods — precisely the headings where the Rauch classification is most economically relevant. By classifying directly at the HS-6 level, PLAID eliminates this source of measurement error and allows researchers to exploit the full granularity of modern trade data. The economic intuition is simple: the way prices form for crude palm oil (HS 151110, exchange-traded) is fundamentally different from that for palm-oil-based cosmetic preparations (HS 340120, differentiated), even though both map to the same SITC heading.

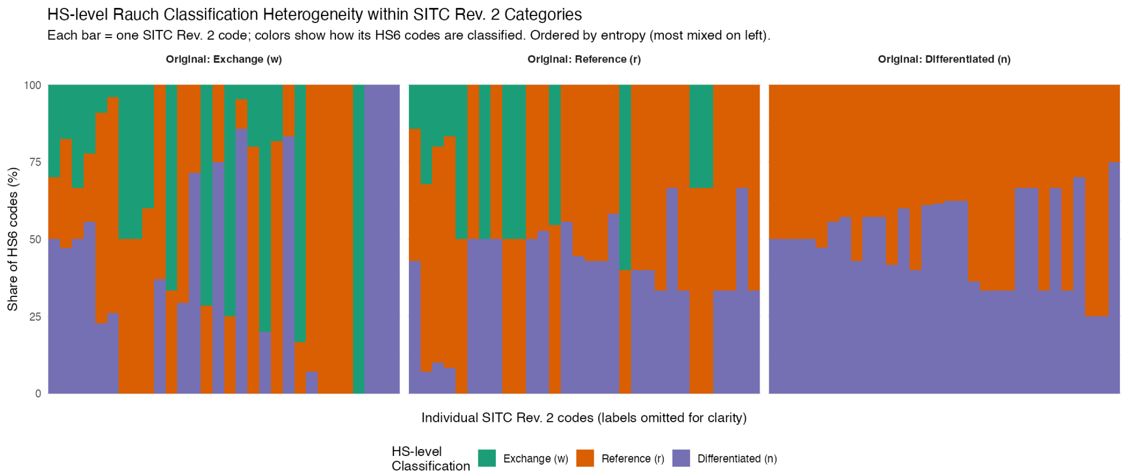

Figure 3: Within-SITC heterogeneity in HS-6 Rauch classifications. Each bar represents a SITC Rev. 2 code, decomposed into the shares of its mapped HS-6 codes classified as organized-exchange (w), reference-priced (r), or differentiated (n). Substantial within-SITC heterogeneity implies that concordance-based approaches mask meaningful variation in price-formation institutions.

Gravity application: networks versus markets. We replicate Rauch's (1999) canonical gravity test — that distance and common language should have differential effects across product categories — using his exact specification (OLS on log bilateral trade, 63-country sample, undirected pairs, zeros dropped) and compare the original SITC-based classification with our LLM-coded SITC classification. Table 3 reports the results. The first three columns reproduce Rauch's (1999, Table 6) published coefficients for the conservative aggregation using 1990 trade data. The last three columns apply the LLM majority-vote classification to the same SITC trade data. The common-language coefficient increases monotonically from organized-exchange to differentiated goods in both — the information-frictions signature that motivates the Rauch classification. Table 10 in the appendix extends the exercise to HS-6 trade data, comparing the PLAID consensus classification with the original Rauch labels mapped through concordances; the PLAID classification produces an even steeper language gradient (0.86, 0.89, 1.16), consistent with finer product granularity sharpening the distinction between categories.

3.2 Broad Economic Categories

Definition. The UN BEC framework groups goods by end use into capital goods (used in production over multiple periods), intermediate goods (transformed or consumed in production), and consumption goods (for final household use). BEC was not constructed

Table 3: Gravity regressions by Rauch category: original versus LLM classification (SITC, 1990)

	Rauch (1999)			LLM (SITC)		
	Org.	Ref.	Dif.	Org.	Ref.	Dif.
$GDP_i \cdot GDP_j$ (log)	0.790^a	0.875^a	0.960^a	0.866^a	0.858^a	0.961^a
	(0.031)	(0.022)	(0.021)	(0.035)	(0.020)	(0.023)
$GDPPC_i \cdot GDPPC_j$ (log)	-0.066^b	0.099^a	0.198^a	-0.153^a	0.047^b	0.217^a
	(0.033)	(0.023)	(0.022)	(0.039)	(0.022)	(0.025)
Distance (log)	-0.701^a	-0.830^a	-0.754^a	-0.963^a	-0.923^a	-0.938^a
	(0.074)	(0.052)	(0.052)	(0.080)	(0.048)	(0.054)
Adjacent	1.223^a	1.016^a	0.945^a	0.533	0.386^c	0.283
	(0.309)	(0.224)	(0.225)	(0.355)	(0.219)	(0.246)
Common language	0.425^a	0.660^a	0.866^a	0.625^a	0.597^a	0.826^a
	(0.153)	(0.108)	(0.107)	(0.193)	(0.115)	(0.129)
EEC/EU	0.201	0.058	0.030	-0.697^c	−0.101	−0.014
	(0.329)	(0.240)	(0.241)	(0.365)	(0.226)	(0.253)
EFTA	−1.148	−0.108	0.150	−0.358	−0.173	−0.109
	(0.498)	(0.362)	(0.365)	(0.649)	(0.402)	(0.451)
n	1603	1724	1804	1523	1677	1687
R^2	0.416	0.668	0.723	0.397	0.681	0.713

OLS. Dependent variable: log average bilateral trade (undirected). Zeros dropped. Rauch (1999) columns reproduce his Table 6 (conservative, 1990 trade, 63-country sample). LLM columns use the same specification with majority-vote classifications from GPT-5 and Claude 3.5 Sonnet applied to SITC Rev. 2 descriptions. Gravity variables from CEPII. Standard errors in parentheses.
^{a} $p < 0.01$, ^{b} $p < 0.05$, ^{c} $p < 0.10$

through a formal classification algorithm. The original version, issued in 1971, was devised as a means of converting SITC-based trade data into System of National Accounts end-use categories (United Nations. Statistical Office, 1971), with UN statistical experts assigning each product heading to a BEC category by reading descriptions and applying a fixed set of conceptual distinctions. The four revisions between 1976 and 2002 primarily conformed BEC to successive SITC and HS editions without reconsidering product-level assignments. The fifth and most recent revision, endorsed in 2016, more substantially restructured the top-level categories, explicitly separated broad economic categories from end-use categories, and extended coverage to services, and was defined natively against HS 2012 (United Nations Statistical Division, 2016). Coverage of HS 2017 and HS 2022 codes therefore continues to rely on correspondence tables rather than fresh product-level assignments. Across all revisions, the assignment of individual HS codes to end-use categories has relied on expert judgment applied to product descriptions rather than any formal algorithm—and the UNSD's own documentation acknowledges that, for borderline products, this judgment cannot always be resolved from descriptions alone (United Nations Statistical Division, 2015).

Our LLM pipeline replicates this assignment task directly on HS descriptions, bypassing the correspondence-table chain entirely. The prompt applies the same conceptual end-use rules used in the original BEC construction—capital for goods used in production over multiple periods, intermediate for inputs consumed or transformed in production, and consumption for goods destined for final household use—with explicit tie-breaker rules that mirror UNSD guidance. Because the method operates on descriptions rather than concordances, it produces internally consistent classifications across all HS revisions since 1992. The ensemble uncertainty measure u_k naturally flags borderline cases in which model disagreement is greatest, providing a direct signal of the classification difficulty that the UNSD itself acknowledges cannot be fully resolved from product text alone.

```
You are a trade-economics specialist. Classify the HS-6 item
by primary end-use (UN BEC framework):

    "capital": machinery, equipment, durable producer goods
    "intermediate": raw materials, parts, components, industrial
        inputs
    "consumption": finished goods for final household use

Tie-breaker: if a good is predominantly used as an input in
further production, classify as intermediate even if it can
also be consumed directly.

Product code: HS 847130
Description:  "Portable automatic data-processing machines,
weighing not more than 10 kg"
chapter_hint: 84 -- Nuclear reactors, boilers, machinery

Return JSON: {"bec": "capital", "reasoning": "...", "confidence
   ": 0.75}
```

The full prompt is reproduced in Appendix B.2.

Validation. We validate the LLM-generated BEC labels against the UN BEC Rev. 5 classification at its native HS vintage, HS 2012 (H4), sourced directly from the official UNSD HS12BEC5 correspondence (United Nations Statistical Division, 2016). This avoids injecting HS concordance noise into the validation: the benchmark and the ensemble both describe the same 5,203 HS 2012 codes with complete BEC assignments. Across these codes the LLM ensemble achieves 91.2% overall agreement with the UNSD benchmark. Per-category agreement is 84.4% for capital, 91.5% for intermediate, and 93.7% for consumption goods (Table 4). The confusion matrix (Figure 4) shows that most remaining disagreement occurs at the capital/intermediate boundary, consistent with the long-standing ambiguity around producer durables that serve both roles.

Cross-revision robustness. Because BEC Rev. 5 is defined natively against HS 2012, measuring agreement at H5 (HS 2017) or H6 (HS 2022) requires propagating the official

Table 4: BEC validation against UN BEC Rev. 5 at HS 2012 (H4)

BEC validation at H4 (native HS 2012)	
N matched	5203
Overall agreement	91.2%
Per-category agreement	
Capital	84.4%
Intermediate	91.5%
Consumption	93.7%

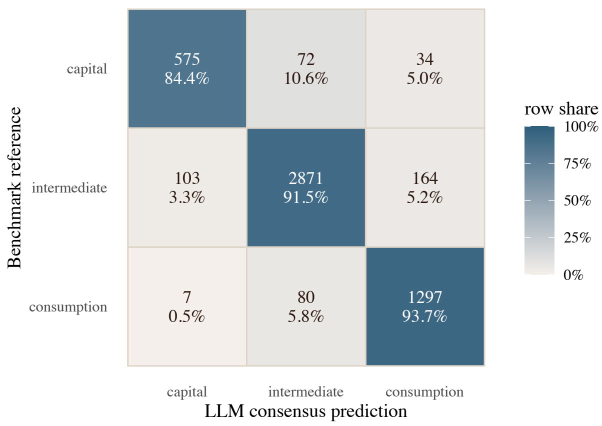

Figure 4: BEC validation at HS 2012 (H4). LLM ensemble vs. UNSD BEC Rev. 5 benchmark on 5,203 HS-6 codes. Overall agreement 91.2%.

assignments through WCO concordances. If the LLM's description-based approach is genuinely robust to HS drift, we should see the H4 agreement levels carry over essentially unchanged. Table 5 confirms this: agreement is 91.1–91.2% at H4, H5, and H6 — flat rather than declining. Put differently, the ensemble's classification of an HS 2017 or HS 2022 code from its description alone matches the UN concordance at the same rate as the native H4 validation, and both constructions are near-equivalent measurements of BEC end-use at the product level.

Table 5: BEC agreement by HS revision (LLM ensemble vs. UNSD concordance)

Revision	**Benchmark**	**N**	**Agreement**
H4	HS 2012 native	5203	91.2%
H5	HS 2017	5385	91.1%
H6	HS 2017 (crosswalked forward)	5240	91.1%

3.3 Perishability & Cold-Chain Requirement

Definition. We classify products on a five-class economic perishability scale based on how quickly the good loses economic value over time, encompassing physical spoilage, chemical instability, regulatory expiry, seasonal/fashion obsolescence, and technological obsolescence. Each product also receives a numeric *economic* half-life estimate in days.

```
Classify the product's economic perishability:
- 1 = Ultra-perishable (1-7 days)
- 2 = Highly perishable (8-60 days)
- 3 = Moderately perishable (61-360 days)
- 4 = Low perishability (1-10 years)
- 5 = Non-perishable (>10 years)
Product: HS 030289, "Fish, fresh or chilled"
Return JSON: {"perishability_class": 1, "half_life_days": 3}
```

The full prompt is reproduced in Appendix B.3.

Validation. We cross-validate against HS chapter-level expectations: agricultural products (chapters 01–24) should skew toward classes 1–3, while metals and machinery toward class 5.

The results confirm strong face validity. Agriculture and food products (chapters 01–24) receive a mean perishability class of 2.59, with 32% of codes classified as ultra- or highly perishable (classes 1–2). Chemicals and plastics (chapters 25–40) average 3.67, textiles and miscellaneous manufactures (chapters 41–67) average 3.57, and metals (chapters 68–83) average 4.77 — with zero codes in classes 1–2. Machinery and transport equipment (chapters 84–97) average 3.87. The ordering Agriculture $<$ Chemicals $<$ Textiles $<$ Machinery $<$ Metals is monotonic and aligns with prior expectations about economic shelf life (Figure 5). Across the four-model ensemble, unanimous agreement on perishability class is 32.9%, but agreement within ± 1 class reaches 92–100%, with Spearman's ρ between 0.71 and 0.82.

Gravity application: adjacency premium. Perishable goods should exhibit a much larger contiguity (shared-border) effect than non-perishable goods, reflecting the "freshness premium" of fast overland transport versus sea freight. We interact perishability class with the contiguity dummy.

Table 6 confirms the hypothesis strongly. Ultra-perishable goods (class 1) face a distance elasticity of -1.54, nearly double the -0.79 to -0.84 range observed for classes 3–5. The contiguity premium is similarly largest for ultra-perishable goods (0.80) versus 0.30–0.49 for other classes, consistent with adjacent countries capturing a "freshness premium" through fast overland transport.

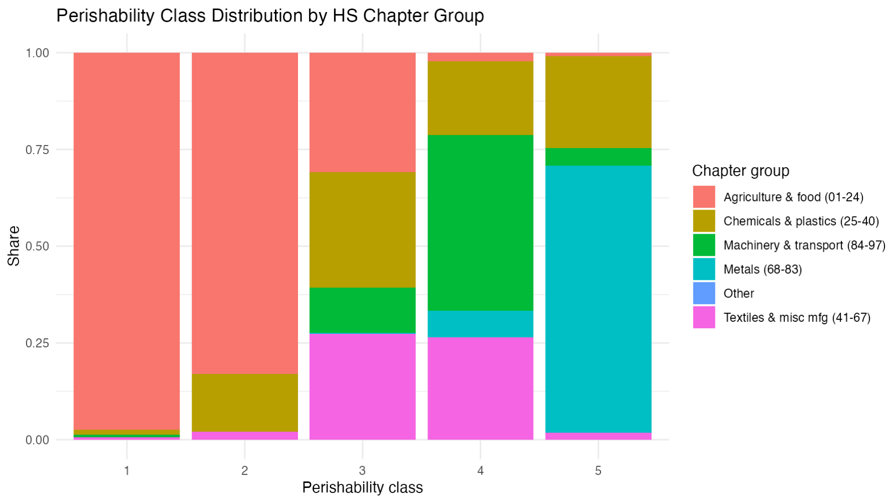

Figure 5: Perishability class distribution by HS chapter group. Stacked shares of perishability classes (1 = ultra-perishable to 5 = non-perishable) within each chapter group.

Table 6: Structural gravity by perishability class (PPML, 2019)

Dependent Variable:	Trade value				
Model:	(1)	(2)	(3)	(4)	(5)
Distance (log)	-1.54^a	-0.803^a	-0.784^a	-0.792^a	-0.840^a
	(0.115)	(0.068)	(0.064)	(0.058)	(0.061)
Contiguity	0.801^a	0.439^a	0.297^b	0.410^a	0.487^a
	(0.153)	(0.126)	(0.118)	(0.114)	(0.142)
Common language	0.095	0.100	0.359^a	0.193^c	0.074
	(0.270)	(0.102)	(0.114)	(0.109)	(0.104)
Colonial tie	-0.372^c	-0.107	0.015	-0.169	0.164
	(0.209)	(0.269)	(0.150)	(0.140)	(0.157)
Fixed effects					
Exporter	Yes	Yes	Yes	Yes	Yes
Importer	Yes	Yes	Yes	Yes	Yes
Observations	10,579	21,606	27,856	28,654	26,970
R^2_{pseudo}	0.86848	0.89801	0.93234	0.94898	0.88609

Clustered (Exporter & Importer) standard-errors in parentheses
Signif. Codes: a: 0.01, b: 0.05, c: 0.1

FEs: origin, destination. Dependent variable: bilateral trade value.

3.4 Hazardous Materials

Definition. We assign a binary flag indicating whether a product is *hazardous*: classified under the Globally Harmonized System (GHS) of Classification and Labelling of Chemicals, or subject to dangerous-goods transport regulations such as the IMDG Code (maritime) and IATA Dangerous Goods Regulations (air). This encompasses explosives, flammable materials, toxic substances, corrosives, radioactive materials, and oxidizers, focusing on the inherent physical-chemical hazard of the typical product fitting the HS-6 description.

```
Classify the product:
- "hazardous": true if GHS-classified or dangerous goods
```

```
  (IMDG / IATA-DGR)
Product: HS 281511, "Sodium hydroxide (caustic soda)"
Return JSON: {"hazardous": true}
```

The full prompt is reproduced in Appendix B.4.

Validation. No product-level benchmark for GHS/IMDG exposure exists at HS-6. Instead, we exploit the strong prior that the hazardous flag should load almost entirely on a small set of chemistry, fuel, and explosives chapters and be near-zero in unambiguously benign chapters. We designate *expected-hazardous* chapters as 27 (mineral fuels), 28 (inorganic chemicals), 29 (organic chemicals), 31 (fertilizers), 36 (explosives and pyrotechnics), and 38 (miscellaneous chemical products), and *expected-non-hazardous* chapters as 41–46 (hides, leather, furs, wood, cork, straw), 61–67 (apparel, headgear, umbrellas, feathers), 94 (furniture and lighting), and 95 (toys, games, and sports equipment). A well-calibrated classifier should produce a high hazardous share in the former and near zero in the latter.

Table 7 confirms the pattern. Across the 726 HS-6 codes in the expected-hazardous chapter group, 80.7% are flagged as hazardous; across the 580 codes in the expected-non-hazardous group, only 0.5%; across all other chapters, 5.4%. The contrast between chapter groups exceeds two orders of magnitude. Figure 6 decomposes the hazardous share by individual chapter within each group: explosives (Chapter 36), organic and inorganic chemicals (29, 28), miscellaneous chemicals (38), and mineral fuels (27) all exceed 70% hazardous; every textile, leather, furniture, and toy chapter in the non-hazardous group sits at zero or a single-digit percentage, with the sole exception of wood (Chapter 44) at around 3% driven by treated-wood preparations.

Table 7: Hazardous flag: HS chapter face validity

Chapter group	N codes	Hazardous share
Expected hazardous (27-38)	726	80.7%
Other	4082	5.4%
Expected non-hazardous	580	0.5%

3.5 Microchip / Semiconductor Content

Definition. A binary flag indicating whether the product contains, embeds, or is a semiconductor, microprocessor, integrated circuit, or microcontroller as a functional component. This covers finished electronics (laptops, smartphones), embedded systems (modern automobiles, medical devices), and the chips themselves.

```
Does this product contain a semiconductor as a functional
   component?
```

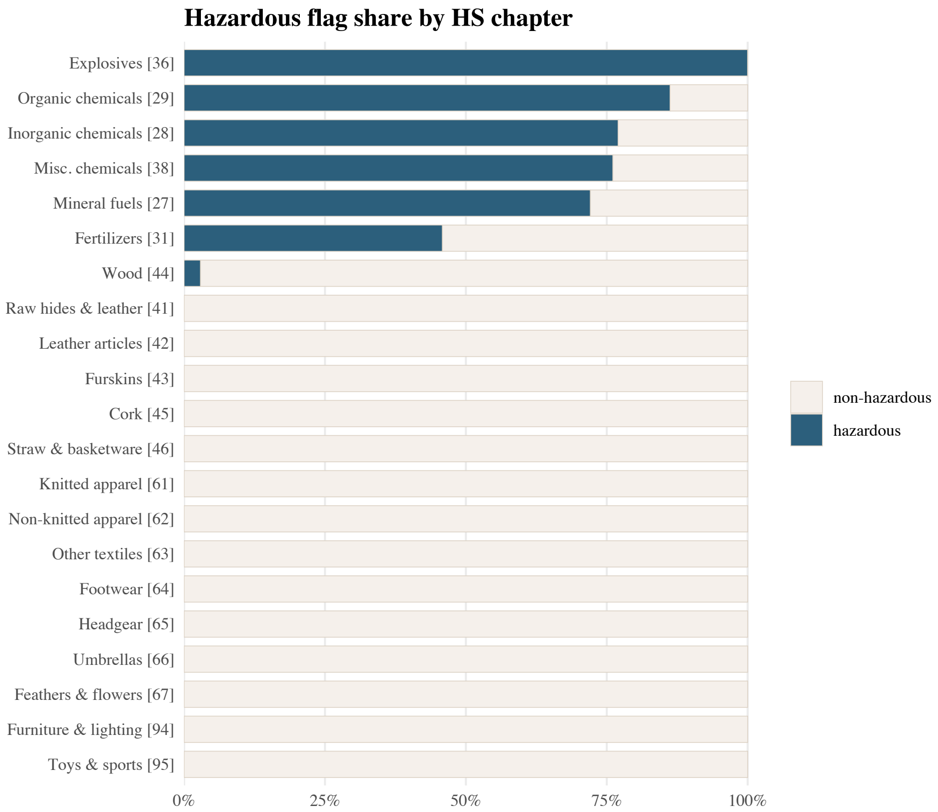

Figure 6: Hazardous-flag share by HS chapter. Expected-hazardous chapters (27, 28, 29, 31, 36, 38) dominate; expected-non-hazardous chapters (41–46, 61–67, 94, 95) are at or near zero.

```
Product: HS 847130, "Portable automatic data-processing machines
    "
Return JSON: {"microchip_content": true}
```

The full prompt is reproduced in Appendix B.5.

Validation. We compare against the OECD/UNCTAD ICT goods definition (2017 revision), which identifies a narrow set of HS-6 codes as dedicated ICT products. Our microchip indicator is deliberately broader: it covers not only dedicated ICT goods but also products with *embedded* semiconductors (vehicles with ECUs, smart appliances, medical devices), so the LLM classification should be a superset of the ICT list. Across 5,243 matched codes the overall agreement is 92.7%, with 71.6% of OECD ICT positives also flagged by the ensemble (Figure 7). Most codes the LLM flags as microchip-containing are outside the narrow OECD ICT list, consistent with the superset interpretation: these are the vehicles, appliances, and instruments that carry embedded chips but are not classified as dedicated ICT goods.

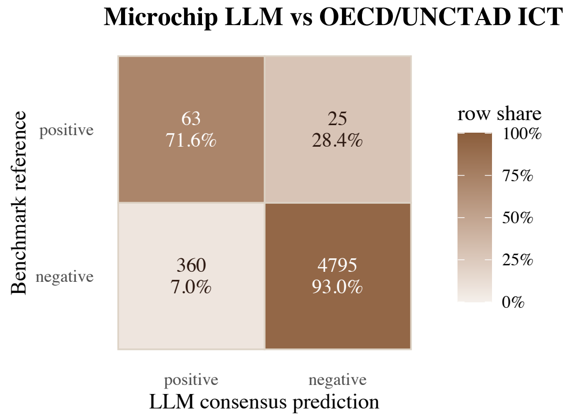

Figure 7: Microchip vs. OECD/UNCTAD ICT. The LLM classification is a superset of the OECD ICT list; the off-diagonal count is dominated by embedded-chip goods that sit outside dedicated ICT.

Chapter face validity. The OECD ICT comparison is informative about coverage of a narrow positive set but is uninformative about specificity over the rest of the HS universe. For that we turn to a chapter-level face-validity test against the prior that microchips should be concentrated in chapters 84 (machinery), 85 (electrical machinery), and 90 (optical, photographic, and medical instruments). Table 8 shows that 56.4% of the 925 HS-6 codes in the expected microchip-bearing chapters are flagged, versus 3.0% across the other 4,463 codes, a nearly twentyfold difference. Figure 8 breaks this down by chapter: optical/medical (Chapter 90) has the highest share at 75%, followed by machinery (Chapter 84) at 57% and electrical machinery (Chapter 85) at 44%. The ordering is consistent with the economic intuition that modern medical and optical instruments are nearly always microprocessor-controlled, while Chapter 85's lower share reflects the large number of passive electrical components (connectors, cables, wire) that do not themselves contain a chip.

Table 8: Microchip flag: HS chapter face validity

Chapter group	N codes	Microchip share
Expected microchip-bearing (84, 85, 90)	925	56.4%
Other chapters	4463	3.0%

3.6 3TG Conflict Minerals

Definition. We flag products that are or contain tin, tantalum, tungsten, or gold — the "3TG" minerals regulated under EU Regulation 2017/821 and US Dodd-Frank Section 1502 due to their association with armed conflict financing. In addition to a binary flag, we

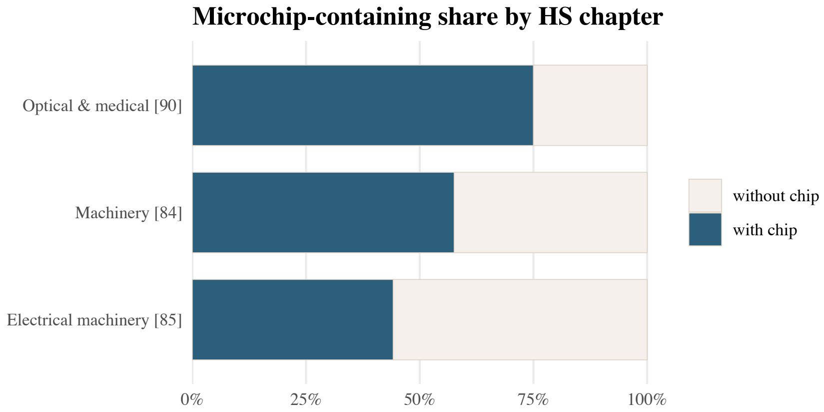

Figure 8: Microchip-containing share by HS chapter. Within the expected microchip-bearing chapters (84, 85, 90), each bar shows the share of HS-6 codes flagged as containing a chip. Optical/medical is highest because modern instruments in that chapter are nearly always microprocessor-controlled.

identify the specific mineral.

```
Does this product contain 3TG conflict minerals?
Product: HS 261590, "Niobium, tantalum or vanadium ores"
Return JSON: {"conflict_mineral": true, "specific_mineral": "
   tantalum"}
```

The full prompt is reproduced in Appendix B.6.

Validation. We compare against EU Regulation 2017/821 Annex I, which lists 22 specific HS-6 codes for regulated 3TG ores, concentrates, and metals. The LLM should match these precisely and, additionally, flag derivative products (alloys and manufactured articles containing 3TG metals).

Of the 20 benchmark codes present in our classification, the ensemble correctly identifies 17 as conflict minerals (85.0% agreement). All codes the ensemble flags as 3TG are true positives. Per-mineral agreement is tin 75.0% (3/4), tantalum 66.7% (2/3), tungsten 88.9% (8/9), and gold 100% (4/4); among true positives, the ensemble correctly identifies the specific mineral in 100% of cases. The three false negatives are codes at the boundary of the regulation's scope (e.g. mixed-oxide headings where 3TG content is partial). The ensemble additionally flags downstream derivative products (alloys, manufactured articles containing 3TG metals) that fall outside the regulation's narrow scope but are relevant for supply-chain analysis. Across the four-model ensemble, unanimous agreement on the conflict-mineral flag reaches 97.6% — the highest of all six indicators.

Chapter face validity. Complementing the narrow EU 2017/821 benchmark, we check that the 3TG flag loads on the chapters where tin, tantalum, tungsten, and gold should concentrate: Chapter 26 (ores, slag, and ash), 71 (pearls, precious stones, and precious metals), 80 (tin and articles thereof), and 81 (other base metals, including tungsten and tantalum). Table 9 and Figure 9 show that 23.8% of codes in these chapters carry the flag, versus only 1.6% elsewhere. Chapter 80 (tin) is essentially fully flagged, Chapter 71 sits near 28% reflecting the share of pearl/diamond sub-headings that do not contain gold, and Chapter 26 is dominated by non-3TG ores such as iron and copper.

Table 9: 3TG flag: HS chapter face validity

Chapter group	N codes	3TG share
Expected 3TG (26, 71, 80, 81)	143	23.8%
Other chapters	5245	1.6%

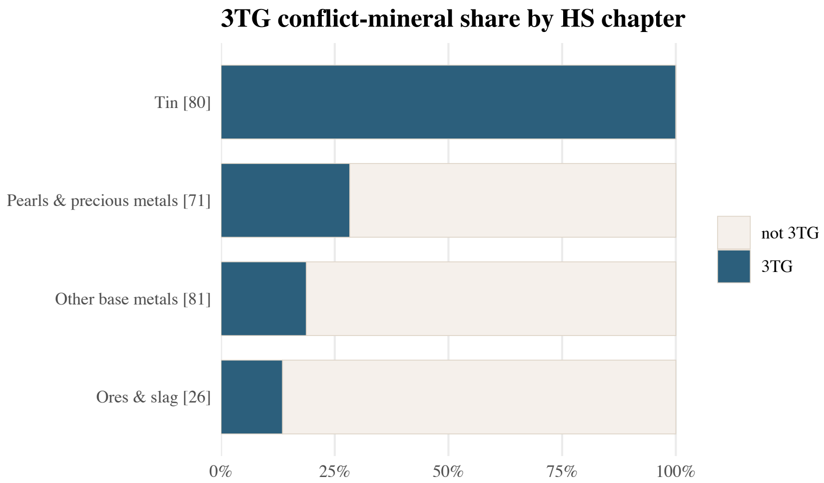

Figure 9: 3TG flag share by HS chapter. Expected 3TG chapters 26, 71, 80, 81 (tin is essentially fully flagged) against all other chapters.

4 The PLAID Database

PLAID is released as a set of 42 gzipped CSV files — one per indicator–revision combination — plus a companion website with an interactive product explorer and a static JSON API.

File structure. Files follow the naming convention `PLAID_v0.1_{indicator}_{revision}.csv.gz`, where `indicator` ∈ {`rauch`, `bec`, `perishability`, `hazmat`, `microchip`, `3tg`} and `revision` ∈ {`H0`, `H1`, ..., `H6`}, corresponding to HS 1992 through HS 2022. Each file contains approximately 5,000–5,600 HS 6-digit codes for the corresponding HS revision.

All files share two common columns: `hs6_code` (zero-padded 6-digit string) and `description` (official product description from the HS nomenclature). Indicator-specific columns are:

Consensus values are majority votes across the four-model ensemble. The share, mean, and

Indicator	Columns	Values
rauch	rauch, rauch_share_w/r/n	w, r, n
bec	bec, bec_share_*	capital, intermediate, consumption
perishability	perishability_class, _mean, _sd, half_life_days_mean, _sd	1–5
hazmat	hazardous (+ _mean, _sd)	TRUE/FALSE
microchip	microchip_content (+ _mean, _sd)	TRUE/FALSE
3tg	conflict_mineral (+ _mean, _sd), specific_mineral	TRUE/FALSE tin/tantalum/tungsten/gold/none

standard deviation columns quantify inter-model agreement and can be used as continuous uncertainty measures in robustness checks.

Website and API. The database is hosted at `https://plaid.julianhinz.com`, which provides an interactive product explorer (searchable and sortable by any indicator), bulk CSV downloads, and a static JSON API. Individual products can be queried at `/api/v0.1/{revision}/{code}.json`; full revision indices at `/api/v0.1/{revision}/index.json`. Machine-readable documentation following the `llms.txt` standard is available at `/llms.txt`.

Versioning. PLAID is versioned (currently v0.1 beta). Each release specifies the date, the ensemble of models used, and HS revision coverage. Future releases will add indicators and expand model coverage.

Replication. All code is available at `https://github.com/julianhinz/PLAID`.

License. PLAID is released under CC BY 4.0. We ask users to cite this paper when using indicators from the database.

5 Conclusion

We have introduced an open database of LLM-generated product-level indicators for international trade research. The current beta release covers six indicators — Rauch classification, BEC end-use, perishability, hazardous materials, microchip content, and 3TG conflict minerals — each validated against an established benchmark or a targeted face-validity exercise, with gravity applications for the two indicators where the economic test is sharpest (Rauch and perishability). PLAID is designed as a living resource: future releases will add indicators for technology intensity, R&D intensity, durability, regulatory sensitivity, and other product attributes.

The broader contribution is methodological. By treating product classification as a text task over official HS descriptions, the LLM-based approach offers three advantages over traditional concordances: (i) automatic coverage of new HS revisions and newly introduced product codes; (ii) scalability to arbitrary indicator definitions without manual coding; and (iii) built-in uncertainty quantification through multi-model ensembles. We hope the open release of data and code will invite replication and critique.

A HS-Level Gravity Regressions

Table 10 extends the gravity replication from Section 3.1 to HS-6 trade data (1995, the earliest year with HS92 bilateral flows). The left panel applies the PLAID four-model consensus classification; the right panel maps Rauch's (1999) original SITC labels to HS92 codes via the WITS concordance, using majority vote where multiple SITC codes map to one HS-6 code.

Table 10: Gravity regressions by Rauch category: PLAID versus concordance-based classification (HS92, 1995)

	PLAID (HS)			Rauch via concordance		
	Org.	Ref.	Dif.	Org.	Ref.	Dif.
$GDP_i \cdot GDP_j$ (log)	1.002^a	0.971^a	1.038^a	0.999^a	1.038^a	1.039^a
	(0.035)	(0.019)	(0.019)	(0.023)	(0.020)	(0.020)
$GDPPC_i \cdot GDPPC_j$ (log)	-0.232^a	0.060^a	0.258^a	0.055^b	0.103^a	0.218^a
	(0.036)	(0.021)	(0.020)	(0.025)	(0.021)	(0.021)
Distance (log)	-1.045^a	-0.883^a	-0.769^a	-0.661^a	-0.895^a	-0.861^a
	(0.079)	(0.047)	(0.046)	(0.055)	(0.048)	(0.048)
Adjacent	0.569	0.479^b	0.474^b	0.740^a	0.451^b	0.382^c
	(0.348)	(0.211)	(0.208)	(0.249)	(0.218)	(0.218)
Common language	0.862^a	0.887^a	1.159^a	1.039^a	1.107^a	1.069^a
	(0.179)	(0.107)	(0.104)	(0.127)	(0.111)	(0.109)
EEC/EU	-0.673^c	-0.249	-0.178	0.297	-0.171	-0.318
	(0.348)	(0.212)	(0.208)	(0.251)	(0.219)	(0.218)
EFTA	0.124	0.035	-0.121	-0.083	-0.019	-0.022
	(0.613)	(0.375)	(0.369)	(0.443)	(0.387)	(0.387)
n	1507	1668	1691	1625	1653	1676
R^2	0.452	0.737	0.798	0.661	0.751	0.776

OLS. Dependent variable: log average bilateral trade (undirected). Zeros dropped. PLAID columns apply the four-model HS-6 consensus to HS92 trade (1995). Concordance columns map Rauch (1999) conservative SITC labels to HS92 via WITS concordance (majority vote where multiple SITC codes map to one HS code) and use the same 1995 HS92 trade. Both use Rauch's 63-country sample. Gravity variables from CEPII. Standard errors in parentheses.
^{a} $p < 0.01$, ^{b} $p < 0.05$, ^{c} $p < 0.10$

B Verbatim Prompts

This appendix reproduces the complete prompts used for each of the six PLAID v0.1 indicators, as they appear in the source build scripts in `code/create_hs_*/02_build_prompts_*.R`. Each prompt is shown with its system-level role and task block, classification guidance, few-shot examples, and the JSON output schema. The input block shown is a representative HS-6 record; at run time this block is replaced per code. For the hazardous-materials prompt, the reproduction has been edited to match the hazardous-only schema delivered by the v0.1 database; the original source script retains the broader language used during the April 2026 run (see Section 2).

B.1 Rauch classification

```
You are a trade-economics specialist.

Task:
Classify each item into three categories with labels w, r, and n
    .

Definitions:
- w = organized-exchange: homogeneous goods with standardized
    contracts and public exchange prices.
- r = reference-priced: not exchange-listed, but widely
    published benchmark prices exist in trade journals or price-
    reporting agencies.
- n = differentiated: no exchange listing and no widely accepted
     benchmark prices; valuation depends on brand, design, and
    specifications.

Decision order:
1) Exchange test → w
2) Reference-price test → r
3) Else → n

Use the most specific described grade/form (not just the broad
    heading).
Tie-breaker: when evidence is strong, prefer the more
    standardized category (w > r > n).
If information is insufficient, make a best guess with lower
    confidence and add a flag.

Cues:
- Exchange (w): exchange contracts such as CBOT/ICE/LME grades (
    for example, No. 2 yellow corn, LME Grade A copper).
- Reference (r): standardized specs with widely published spot
    or benchmark assessments (polymers, fertilizers, common steel
     forms).
- Differentiation (n): brand or model identifiers; consumer
    goods, apparel, machinery, bespoke parts.

Few-shot examples (format to imitate):

```

```
Example 1
Input → code_system: HS, code: 100590, description: "Maize (corn
    )"
Output → {
  "code_system": "HS",
  "code": "100590",
  "short_description": "Maize (corn)",
  "rauch_category": "w",
  "reasoning": "Maize has standardized exchange contracts (e.g.,
      CBOT) with public prices, so it is organized-exchange.",
  "evidence_type": "exchange_listing",
  "confidence": 0.94,
  "flags": []
}

Example 2
Input → code_system: HS, code: 390110, description: "
    Polyethylene with density < 0.94"
Output → {
  "code_system": "HS",
  "code": "390110",
  "short_description": "Polyethylene (<0.94 g/cc)",
  "rauch_category": "r",
  "reasoning": "Polyethylene trades on standardized specs with
      widely published benchmark prices but is not exchange-
      listed, so it is reference-priced.",
  "evidence_type": "reference_prices",
  "confidence": 0.90,
  "flags": []
}

Example 3
Input → code_system: HS, code: 847130, description: "Portable
    automatic data processing machines (laptops)"
Output → {
  "code_system": "HS",
  "code": "847130",
  "short_description": "Laptops",
  "rauch_category": "n",
  "reasoning": "Laptops are branded and differentiated consumer
      goods with no exchange or benchmark prices, so they are
      differentiated.",
  "evidence_type": "differentiation_markers",
  "confidence": 0.88,
  "flags": []
```

```
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
Return JSON only (one object). Use this schema (types/allowed values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"rauch_category": "w|r|n",
"reasoning": "1-2 sentences citing exchange/reference/differentiation cues",
"evidence_type": "exchange_listing|reference_prices|differentiation_markers|heuristic_guess",
"confidence": "number 0.0-1.0",
"flags": ["optional notes like 'ambiguous_grade',' insufficient_detail'"]
}
```

B.2 Broad Economic Categories

```
You are a trade-economics specialist.

Task:
For the given HS-6 item, classify its end-use category according to the UN Broad Economic Categories (BEC).

Categories:
- capital: Goods used in production over multiple periods (machinery, equipment, tools, vehicles used in production).
- intermediate: Goods that are transformed or consumed in the production process (raw materials, parts, components, fuels, semi-finished products).
- consumption: Goods destined for final household use (food for household consumption, clothing, consumer electronics, household appliances).

```

```
Important:
- Focus on the PRIMARY end-use of the typical traded form of the product.
- Some goods have dual uses; classify according to the predominant use in international trade.
- Industrial machinery and equipment → capital.
- Raw materials, ingredients, intermediate inputs → intermediate.
- Finished consumer goods for households → consumption.

Outputs required:
1) bec: one of capital, intermediate, consumption
2) a short reason grounded in end-use, production chain position, and trade patterns.

Classification guidance:
- capital: Equipment with productive life spanning multiple periods.
  Examples: turbines, lathes, tractors (agricultural machinery), aircraft, ships, construction equipment.
- intermediate: Inputs consumed or embodied in production, traded between firms.
  Examples: crude oil, iron ore, cotton fibers, automotive parts, electronic components, chemicals for industry.
- consumption: Final goods purchased by households for direct use.
  Examples: packaged food, beverages, apparel, footwear, consumer electronics, household appliances, pharmaceuticals for retail.

Tie-breakers:
- If a good is predominantly used as an input in further production, classify as intermediate even if it can also be consumed directly.
- If a good is a finished product sold primarily to households, classify as consumption even if businesses also buy it.
- Capital goods that are also consumer durables (e.g. personal computers) → consumption if primarily household use, capital if primarily business/industrial.

Return JSON only. No extra text outside JSON.

Few-shot examples (format to imitate):

Example 1
```

```
Input → code_system: HS, code: 840734, description: "Spark-ignition reciprocating piston engines of a cylinder capacity exceeding 1,000 cc"
Output → {
  "code_system": "HS",
  "code": "840734",
  "short_description": "Engine parts for vehicles",
  "bec": "intermediate",
  "reasoning": "Internal combustion engines are components assembled into vehicles; they enter further production rather than being consumed directly by households.",
  "confidence": 0.90
}

Example 2
Input → code_system: HS, code: 721391, description: "Bars and rods of iron or non-alloy steel, hot-rolled, in irregularly wound coils"
Output → {
  "code_system": "HS",
  "code": "721391",
  "short_description": "Hot-rolled steel in coils",
  "bec": "intermediate",
  "reasoning": "Hot-rolled steel is a semi-finished input used in manufacturing; it is processed further before reaching end consumers.",
  "confidence": 0.95
}

Example 3
Input → code_system: HS, code: 610910, description: "T-shirts, singlets and other vests, of cotton"
Output → {
  "code_system": "HS",
  "code": "610910",
  "short_description": "Cotton T-shirts",
  "bec": "consumption",
  "reasoning": "Finished cotton garments are final consumer goods purchased directly by households.",
  "confidence": 0.95
}

Example 4
Input → code_system: HS, code: 845010, description: "Household-type washing machines, each of a dry linen capacity not
```

```
exceeding 10 kg"
Output → {
  "code_system": "HS",
  "code": "845010",
  "short_description": "Household washing machines",
  "bec": "consumption",
  "reasoning": "Household washing machines are consumer durables purchased by households for personal use, not for productive investment.",
  "confidence": 0.90
}

Example 5
Input → code_system: HS, code: 847130, description: "Portable automatic data processing machines, weighing not more than 10 kg"
Output → {
  "code_system": "HS",
  "code": "847130",
  "short_description": "Laptops and portable computers",
  "bec": "capital",
  "reasoning": "Portable computers are primarily used as productive capital equipment in offices and businesses, though they also serve household uses.",
  "confidence": 0.75
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
Return JSON only (one object). Use this schema (types/allowed values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"bec": "capital|intermediate|consumption",
"reasoning": "1-2 sentences citing end-use, trade chain position",
"confidence": "number 0.0-1.0"
```

```
}
```

B.3 Perishability

```
You are a trade-economics specialist.

Task:
For the given HS-6 item, classify its economic perishability and
    estimate an approximate value half-life.

Key concept (economic perishability):
How quickly the good loses economic value over time due purely
    to time passing,
even under standard commercial storage and handling for
    internationally traded goods.

This includes (when relevant):
- Physical spoilage or biological decay (food, flowers).
- Chemical instability or sterility loss (some pharmaceuticals/
    biologics).
- Regulatory expiry or shelf-life constraints.
- Seasonal/fashion obsolescence (apparel, seasonal items).
- Technological/model obsolescence (electronics).

Important:
- Do NOT equate perishability with 'food' or HS chapter.
- Assume the typical traded form consistent with the HS-6
    description (e.g., fresh vs frozen; live vs processed).
- All HS goods are storable in principle; your job is to assess
    how costly time is for this product.

Outputs required:
1) perishability_class: integer in {1,2,3,4,5}
2) half_life_days: a single numeric estimate of the value half-
    life in DAYS (not hours),
3) a short reason grounded in shelf-life / obsolescence / time
    sensitivity cues.

Perishability classes (anchor ranges by value half-life in DAYS)
    :
- 1 = Ultra-perishable: 1-7 days (time sensitivity dominates; a
    few days can destroy value).
- 2 = Highly perishable: 8-60 days (days-weeks matter materially
    ; cold chain / expiry common).
```

```
- 3 = Moderately perishable: 61-360 days (months; depreciation/obsolescence within a year).
- 4 = Low perishability: 361-3600 days (1-10 years; slow depreciation).
- 5 = Non-perishable: >3600 days (>10 years; time largely irrelevant).

How to estimate half_life_days:
- Provide a best-guess point estimate consistent with the class anchors.
- Think 'economic value half-life': time until the typical market value is about 50% lower due to time alone.
- For physical perishables: use typical commercial shelf life under proper storage (e.g., cold chain).
- For fashion/tech: use typical product-cycle depreciation (model turnover, seasonal markdowns).

Tie-breakers:
- If the description suggests strict freshness/cold-chain or very short shelf life, move toward lower class (more perishable).
- If the good is a durable input (metals, ores, basic chemicals) or a long-lived capital good, move toward higher class.
- If insufficient detail (e.g., unclear fresh vs processed), choose class 3 with lower confidence and add a flag.

Return JSON only. No extra text outside JSON.

Few-shot examples (format to imitate):

Example 1
Input → code_system: HS, code: 030289, description: "Fish, fresh or chilled (excluding fillets)"
Output → {
  "code_system": "HS",
  "code": "030289",
  "short_description": "Fresh or chilled fish",
  "perishability_class": 1,
  "half_life_days": 3,
  "reasoning": "Fresh/chilled fish loses market value within days even under cold chain; border delays are very costly.",
  "evidence_type": "physical_spoilage",
  "confidence": 0.90,
  "flags": []
```

```
}

Example 2
Input → code_system: HS, code: 610910, description: "T-shirts of cotton"
Output → {
  "code_system": "HS",
  "code": "610910",
  "short_description": "Cotton T-shirts",
  "perishability_class": 3,
  "half_life_days": 180,
  "reasoning": "Apparel is economically perishable via seasonality/fashion and retail markdown cycles; value decays over months.",
  "evidence_type": "seasonal_fashion_obsolescence",
  "confidence": 0.75,
  "flags": []
}

Example 3
Input → code_system: HS, code: 720711, description: "Semi-finished products of iron or non-alloy steel"
Output → {
  "code_system": "HS",
  "code": "720711",
  "short_description": "Semi-finished steel products",
  "perishability_class": 5,
  "half_life_days": 20000,
  "reasoning": "Basic steel forms do not become obsolete quickly and can be stored for years with minimal economic decay; time is largely irrelevant.",
  "evidence_type": "durable_storable_input",
  "confidence": 0.85,
  "flags": []
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
```

```
Return JSON only (one object). Use this schema (types/allowed values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"perishability_class": "integer 1-5",
"half_life_days": "number (single point estimate in days)",
"reasoning": "1-2 sentences citing spoilage/expiry/seasonality/obsolescence cues",
"evidence_type": "physical_spoilage|chemical_instability|expiry_regulatory|seasonal_fashion_obsolescence|tech_obsolescence|durable_storable_input|heuristic_guess",
"confidence": "number 0.0-1.0",
"flags": ["optional notes like 'ambiguous_form','unclear_fresh_vs_processed','insufficient_detail'"]
}
```

B.4 Hazardous materials

```
You are an expert on hazardous materials classification.

Task:
For the given HS-6 item, classify whether it is hazardous.

Key concept:

hazardous: The product is classified under the Globally Harmonized System (GHS) of Classification
and Labelling of Chemicals, or is subject to dangerous goods transport regulations (IMDG, IATA-DGR).
This includes explosives, flammable materials, toxic substances, corrosives, radioactive materials,
and oxidizers. Focus on the inherent physical and chemical hazard of the product itself.

Important:
- Base the classification on the TYPICAL product fitting the HS-6 description,
  not on rare or exotic uses.

Classification guidance:

For hazardous:
```

```
- YES if: the product is a chemical that is toxic, corrosive, explosive, or flammable at
  standard conditions; radioactive material; oxidizing agent; compressed/liquefied gas.
- NO if: the product is an inert solid, fabric, food, wood, or other material without
  inherent chemical/physical hazard properties.

Return JSON only. No extra text outside JSON.

Few-shot examples (format to imitate):

Example 1
Input -> code_system: HS, code: 281511, description: "Sodium hydroxide (caustic soda), solid"
Output -> {
  "code_system": "HS",
  "code": "281511",
  "short_description": "Sodium hydroxide (caustic soda), solid",
  "hazardous": true,
  "reasoning": "Sodium hydroxide is a corrosive substance classified under GHS and subject to IMDG dangerous goods regulations.",
  "confidence": 0.95
}

Example 2
Input -> code_system: HS, code: 520100, description: "Cotton, not carded or combed"
Output -> {
  "code_system": "HS",
  "code": "520100",
  "short_description": "Cotton, not carded or combed",
  "hazardous": false,
  "reasoning": "Raw cotton is an agricultural commodity with no inherent chemical hazard.",
  "confidence": 0.98
}

Example 3
Input -> code_system: HS, code: 284440, description: "Radioactive chemical elements and isotopes"
Output -> {
  "code_system": "HS",
  "code": "284440",
```

```
  "short_description": "Radioactive chemical elements and isotopes",
  "hazardous": true,
  "reasoning": "Radioactive elements pose radiation hazards classified under GHS and IMDG.",
  "confidence": 0.97
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
Return JSON only (one object). Use this schema (types/allowed values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"hazardous": true|false,
"reasoning": "1-2 sentences grounded in specific regulatory frameworks or hazard properties",
"confidence": "number 0.0-1.0"
}
```

B.5 Microchip content

```
You are an expert on electronics, semiconductor technology, and product composition.

Task:
For the given HS-6 item, determine whether the product contains, embeds, or IS a
semiconductor, microprocessor, integrated circuit, or microcontroller as a functional component.

Key concept (microchip_content):
A product has microchip_content = true if:
- It IS a semiconductor component (integrated circuits, processors, microcontrollers, memory chips).
```

```
- It is a finished electronic device that functions via embedded semiconductors (laptops, smartphones, TVs, cameras).
- It is a product with embedded control electronics as a core functional part (motor vehicles with ECUs,
  medical devices with microcontrollers, industrial machinery with PLCs, smart appliances).

A product has microchip_content = false if:
- It contains no semiconductor components as a functional element (raw materials, basic commodities).
- It is purely mechanical, chemical, biological, or textile with no embedded electronics.
- Any electronic connection is incidental rather than functional (e.g., basic electric motors without control circuits).

Important:
- Focus on the typical traded form described in the HS-6 code.
- When a product category spans both chip-containing and non-chip variants, assess the dominant form.
- 'Contains a wire' or 'uses electricity' is NOT sufficient; the product must contain a semiconductor device.

Classification guidance:
- TRUE examples: computers, phones, televisions, cameras, vehicles (post-1980s with ECUs),
  aircraft avionics, medical imaging devices, industrial robots, circuit boards, ICs, memory.
- FALSE examples: raw lumber, steel, cotton textiles, fresh food, furniture, basic hand tools,
  simple electric motors without control circuits, basic light bulbs, wire and cable.
- BORDERLINE: For products where some variants contain chips and others do not, lean toward
  true if the majority of internationally traded units at this HS-6 level contain semiconductors.

Return JSON only. No extra text outside JSON.

Few-shot examples (format to imitate):

Example 1
Input → code_system: HS, code: 847130, description: "Portable automatic data-processing machines"
Output → {
  "code_system": "HS",
```

```
  "code": "847130",
  "short_description": "Portable computers (laptops)",
  "microchip_content": true,
  "reasoning": "Laptops are built around processors, memory chips, and other integrated circuits that are central to their function.",
  "confidence": 0.99
}

Example 2
Input → code_system: HS, code: 870323, description: "Motor vehicles, spark-ignition, 1500-3000cc"
Output → {
  "code_system": "HS",
  "code": "870323",
  "short_description": "Passenger cars 1500-3000cc",
  "microchip_content": true,
  "reasoning": "Modern passenger vehicles contain multiple ECUs and microcontrollers for engine management, safety systems, and infotainment.",
  "confidence": 0.97
}

Example 3
Input → code_system: HS, code: 440710, description: "Lumber, coniferous, sawn or chipped lengthwise"
Output → {
  "code_system": "HS",
  "code": "440710",
  "short_description": "Coniferous sawn lumber",
  "microchip_content": false,
  "reasoning": "Sawn lumber is a raw wood product with no electronic components whatsoever.",
  "confidence": 0.99
}

Example 4
Input → code_system: HS, code: 854231, description: "Electronic integrated circuits: processors and controllers"
Output → {
  "code_system": "HS",
  "code": "854231",
  "short_description": "Processors and microcontrollers (ICs)",
  "microchip_content": true,
```

```
  "reasoning": "This HS code IS the semiconductor component
      itself -- processors and microcontrollers are integrated
      circuits by definition.",
  "confidence": 1.00
}

Example 5
Input → code_system: HS, code: 610910, description: "T-shirts of
     cotton"
Output → {
  "code_system": "HS",
  "code": "610910",
  "short_description": "Cotton T-shirts",
  "microchip_content": false,
  "reasoning": "Cotton T-shirts are textile products with no
      semiconductor or electronic components.",
  "confidence": 0.99
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
Return JSON only (one object). Use this schema (types/allowed
   values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"microchip_content": "true or false (boolean)",
"reasoning": "1-2 sentences explaining whether a semiconductor
   is a functional component",
"confidence": "number 0.0-1.0"
}
```

B.6 3TG conflict minerals

```
You are an expert on mineral supply chains, conflict minerals
   regulation, and product composition.

```

```
Task:
For the given HS-6 item, determine whether the product IS or CONTAINS tin, tantalum, tungsten, or gold
(collectively known as 3TG or conflict minerals), and if so, identify which specific mineral it relates to.

Key definitions:
- conflict_mineral: true if the product is, contains, or is primarily derived from tin, tantalum,
  tungsten, or gold. This includes ores, concentrates, refined metals, alloys, and manufactured
  products where these minerals are a defining material input.
  Regulated under EU Regulation 2017/821 and US Dodd-Frank Act Section 1502.
- specific_mineral: which of the four 3TG minerals the product primarily relates to.
  Values: 'tin', 'tantalum', 'tungsten', 'gold', or 'none' if not a conflict mineral.
  If a product could relate to multiple 3TG minerals, pick the primary or most prominent one.

Important:
- Focus on whether the HS-6 description implies the presence or use of a 3TG mineral as a
  key input or primary material, not just an incidental trace amount.
- Generic categories (e.g., 'electrical machinery') should be classified false unless the
  description specifically identifies a 3TG mineral.
- Include downstream products where a 3TG mineral is definitionally present (e.g., tin-plated
  steel, tungsten carbide tools, gold jewellery).
- Do NOT classify products merely because they might coincidentally use a 3TG mineral in production.

Coverage guidance:
- Tin: includes tin ores, tin metal, tin alloys (solder, pewter, bronze), tin compounds, tin-plated products.
- Tantalum: includes tantalite ores, tantalum metal, tantalum capacitors, tantalum carbide.
- Tungsten: includes wolframite/scheelite ores, tungsten metal, tungsten carbide, tungsten filaments, tungsten alloys.
- Gold: includes gold ores, gold metal (monetary and non-monetary), gold alloys, gold jewellery, gold-plated items,
  gold compounds.
```

```

When to classify as true:
- Ores and concentrates containing a 3TG mineral.
- Pure metal forms (unwrought, wrought, powder, waste/scrap).
- Alloys where a 3TG mineral is a named or definitional component.
- Manufactured articles where the product description explicitly identifies a 3TG material.
- Compounds and chemical preparations of a 3TG element.

When to classify as false:
- Generic product categories with no specific mention of a 3TG mineral.
- Products from other minerals or materials (iron, copper, aluminium, etc.).
- Agricultural goods, textiles, plastics, and other clearly non-mineral products.

Return JSON only. No extra text outside JSON.

Few-shot examples (format to imitate):

Example 1
Input → code_system: HS, code: 261590, description: "Niobium, tantalum or vanadium ores and concentrates"
Output → {
  "code_system": "HS",
  "code": "261590",
  "short_description": "Tantalum ores and concentrates",
  "conflict_mineral": true,
  "specific_mineral": "tantalum",
  "reasoning": "Tantalite ore is the primary source of tantalum, a regulated 3TG conflict mineral under EU 2017/821 and US Dodd-Frank Section 1502.",
  "confidence": 0.95
}

Example 2
Input → code_system: HS, code: 800110, description: "Tin, not alloyed, unwrought"
Output → {
  "code_system": "HS",
  "code": "800110",
  "short_description": "Unwrought tin, not alloyed",
  "conflict_mineral": true,
```

```
  "specific_mineral": "tin",
  "reasoning": "Pure unwrought tin is a direct form of a 3TG conflict mineral regulated under EU 2017/821 and US Dodd-Frank Section 1502.",
  "confidence": 0.99
}

Example 3
Input → code_system: HS, code: 710812, description: "Gold, non-monetary, unwrought (excl. powder)"
Output → {
  "code_system": "HS",
  "code": "710812",
  "short_description": "Non-monetary unwrought gold",
  "conflict_mineral": true,
  "specific_mineral": "gold",
  "reasoning": "Non-monetary unwrought gold is a direct form of the 3TG conflict mineral gold, regulated under EU 2017/821 and US Dodd-Frank Section 1502.",
  "confidence": 0.99
}

Example 4
Input → code_system: HS, code: 261100, description: "Tungsten ores and concentrates"
Output → {
  "code_system": "HS",
  "code": "261100",
  "short_description": "Tungsten ores and concentrates",
  "conflict_mineral": true,
  "specific_mineral": "tungsten",
  "reasoning": "Tungsten ores (wolframite, scheelite) are the primary source of tungsten, a regulated 3TG conflict mineral under EU 2017/821 and US Dodd-Frank Section 1502.",
  "confidence": 0.99
}

Example 5
Input → code_system: HS, code: 520100, description: "Cotton, not carded or combed"
Output → {
  "code_system": "HS",
  "code": "520100",
  "short_description": "Raw cotton, not carded or combed",
  "conflict_mineral": false,
```

```
  "specific_mineral": "none",
  "reasoning": "Raw cotton is an agricultural fibre product with no connection to tin, tantalum, tungsten, or gold.",
  "confidence": 0.99
}

---
INPUT
code_system: HS
code: 100590
description: Maize (corn)
chapter_hint: 10 -- Cereals

OUTPUT
Return JSON only (one object). Use this schema (types/allowed values shown as strings):
{
"code_system": "HS",
"code": "string (6 digits)",
"short_description": "string",
"conflict_mineral": "true or false (boolean)",
"specific_mineral": "tin | tantalum | tungsten | gold | none",
"reasoning": "1-2 sentences explaining the classification",
"confidence": "number 0.0-1.0"
}
```

References

Djankov, Simeon, Caroline Freund, and Cong S. Pham, “Trading on Time,” Technical Report 3909, World Bank 2006.

—, —, and —, “Trading on Time,” *The Review of Economics and Statistics*, 2010, *92* (1), 166–173.

Gaulier, Guillaume and Soledad Zignago, “BACI: International Trade Database at the Product-Level. The 1994–2007 Version,” Working Paper 2010-23, CEPII 2010.

Hummels, David L. and Georg Schaur, “Time as a Trade Barrier,” *American Economic Review*, 2013, *103* (7), 2935–2959.

Rauch, James E., “Networks versus markets in international trade,” *Journal of International Economics*, 1999, *48* (1), 7–35.

Stewart, Ian J., "EU Trade in Dual-Use Items with Conflict-Affected Regions," Technical Report, European Parliament, Directorate-General for External Policies of the Union 2026.

United Nations Statistical Division, "Classification by Broad Economic Categories, Rev. 5 — Explanatory Notes," Technical Report, United Nations 2015. ST/ESA/STAT/SER.M/53/Rev.5.

_ , "Correlation and Correspondence Tables between BEC Rev. 5 and HS 2012/HS 2017," Technical Report, United Nations 2016.

United Nations. Statistical Office, *Classification by Broad Economic Categories*, Vol. Series M, No. 53 of *Statistical Papers*, New York: United Nations, 1971. ST/STAT/SER.M/53; United Nations publication, Sales No. E.71.XVII.12.

Versino, Cristina and Giacomo Cojazzi, "Estimation of Trade Flows Related to Export-Controlled Dual-Use Items: Data Sources, Methodology and Tools," Technical Report EUR 25214 EN, Joint Research Centre, European Commission 2012.