Which Industries Are Actually the Defence Industries? AI Meets IO

Topic: Input-Output Analysis: Industrial policies (2)
Author: Esteban FERNANDEZ-VAZQUEZ
Co-Authors: Andre Carrascal-Incera

The question of which industries constitute the defence industrial base has gained renewed urgency in the context of the war in Ukraine and the European Union's Readiness 2030 strategy. Yet despite its policy relevance, the defence industry resists easy definition within standard statistical frameworks. Official classifications such as NACE do not include a dedicated defence sector: military production is distributed across a range of industries â€” from fabricated metals and electronics to aerospace and specialised machinery â€” whose contribution to defence output is neither exclusive nor uniform across countries. This paper addresses this identification problem directly, asking: which industries are actually the defence industries, and does the answer differ across EU member states?

We propose a two-step methodology that bridges natural language processing and input-output analysis to answer this question empirically. In the first step, we compute semantic similarity scores between the 22 categories of the Wassenaar Arrangement Munitions List (ML) â€” the internationally recognised catalogue of controlled military equipment â€” and the product categories of the Classification of Products by Activity (CPA 2.1) used in Eurostat's National supply-use tables for a set of EU countries. Similarity scores are computed using sentence embeddings from the all-MiniLM-L6-v2 transformer model, which encodes both ML category descriptions and CPA product descriptions as high-dimensional vectors and measures their proximity via cosine similarity. This produces a matrix capturing how semantically related each type of military equipment is to each type of industrial product, regardless of country.

In the second step, this similarity matrix is combined with country-specific supply tables from the studied dataset (Eurostat).The key methodological insight is that cross-country heterogeneity in ML-to-NACE linkages arises entirely from differences in supply table structure, not from differences in the semantic similarity scores, which remain fixed. A country with a diversified manufacturing base will spread defence-relevant production across more industries than a specialised economy. If supply tables were purely diagonal â€” each product produced by a single characteristic industry â€” the approach would collapse to a direct ML-to-NACE mapping, identical across countries and equivalent to a simple relabelling exercise. The off-diagonal entries in supply tables are therefore not a complication but the source of identification.

Data requirements are fully met by publicly available sources: the Wassenaar Arrangement Munitions List (2024), the official CPA 2.1 classification documentation, and Eurostat's national supply-use tables. The methodology is transparent and reproducible, relying on open-source sentence-transformer models and standard IO datasets accessible to any researcher. The framework contributes to the IO literature in two ways. First, it offers a principled, data-driven method for constructing defence industry aggregates that respects the distributed nature of military production across NACE sectors. Second, it generates country-specific industry profiles that can serve as inputs to regional analysis, structural decomposition, or multiplier studies focused on defence investment scenarios â€” directly relevant to current EU industrial policy debates.