Measuring Industrial Policy
A Text-Based Approach*

Réka Juhász (UBC, NBER & CEPR) Nathan Lane (Oxford, CESifo) Emily Oehlsen (Oxford University) Verónica C. Pérez (Boston University)

First draft: August, 2022
This draft: September, 2025

Abstract

Since the 18th century, policymakers have debated the merits of industrial policy (IP). Yet, economists lack basic facts about its use due to measurement challenges. We propose a new approach to IP measurement based on information contained in policy text. We show how off-the-shelf supervised machine learning tools can be used to categorize industrial policies at scale. Using this approach, we validate long-standing concerns with earlier measurement approaches that conflate IP with other types of policy. We apply our methodology to a global database of commercial policy descriptions, and provide a first look at IP use at the country, industry, and year levels (2010-2022). The new data on IP suggest that i) IP is on the rise; ii) modern IP tends to use subsidies and export promotion measures as opposed to tariffs; iii) rich countries heavily dominate IP use; iv) IP tends to target sectors with an established comparative advantage, particularly in high-income countries.

JEL: O25,L52,C38

Keywords: Industrial Policy, Supervised Machine Learning, Natural Language Processing

*Previously titled "The Who, What, When, and How of Industrial Policy: A Text-Based Approach."
We thank Elliott Ash, David Byrne, Chiara Criscuolo, John Andrew Fitzgerald, Jason Garred, Penny Goldberg, Douglas Gollin, Amy Handlan, Gordon Hanson, Tarek Hassan, Giammario Impullitti, Myrto Kalouptsidi, Guy Lalanne, Colin McAuliffe, Phillip McCalman, Suresh Naidu, Emi Nakamura, Ashwin Narayan, Gaurav Nayyar, Nina Pavcnik, Tristan Reed, Dani Rodrik, Francesco Squintani, Chris Woodruff, and seminar audiences at American University, Applied Machine Learning, Economics, and Data Science (AMLEDS) Workshop, Berkeley, Brown, Chapman University, Columbia, Council of Economic Advisors (U.S.), Duke University, Econometric Society Summer Meetings, ETH Zurich, ETH Zurich–Monash–Warwick Textual Data Conference, Federal Reserve Bank of Atlanta Firm Dynamics and the Macroeconomy Conference, Harvard, the IGC-Stanford Firms, Development and Trade Conference, the International Monetary Fund, Michigan, MIT Sloan, National University of Singapore, NBER Summer Institute, New Thinking in Industrial, Innovation and Technology Policy Conference at Columbia, Pakistan Institute of Development Economics, STEG-CEPR-IGC Conference, UC-Irvine, University of Melbourne, University of Miami, University of Washington Comparative Politics, Warwick, the World Bank, and Yale for helpful comments and suggestions. We benefited from conversations with Gerard DiPippo, Ilaria Mazzocco, and scholars at the CSIS; the OECD Productivity, Innovation and Entrepreneurship team, and the UNIDO Research and Industrial Policy Advice Division. We especially thank Johannes Fritz for helpful discussions about the *Global Trade Alert* project, and the GTA team for sharing their data. Ida Brzezinska, Caran Niousha Etemad, Lottie Field, Mikhael Gaster, Diti Jain, Max Marczinek, Esha Vaze, Sarah Wappel, and Megan Yamoah provided outstanding research support. We thank the team of Columbia, Oxford, and UBC annotators for their excellent assistance. We gratefully acknowledge funding from ISERP, STEG/CEPR, SSHRC, and the Alfred P. Sloan Foundation. Contacts: Réka Juhász (reka.juhasz@ubc.ca) and Nathan Lane (nathaniel.lane@economics.ox.ac.uk).

1. Introduction

Since the rise of modern capitalism, social scientists have debated the role of industrial policy—intentional state intervention aimed at transforming the composition of economic activity (List, 1856; Taussig, 1914; Chang, 2002; Rodrik, 2008). Despite these longstanding debates, empirical research on industrial policy (IP) remains incomplete (Harrison and Rodríguez-Clare, 2010). Issues of measurement (Barwick, Kalouptsidi and Bin Zahur, 2024)[p.64] and the lack of systematic data have been significant hurdles to understanding policy (e.g., World Trade Organization (2006)). An enduring challenge is that the same policy instrument (for example, a tariff) can serve both as industrial policy and as a tool for other policy objectives (Harrison and Rodríguez-Clare, 2010; Lane, 2021). As industrial policy experiences a global resurgence, addressing these issues is crucial for understanding these policies.

In this paper, we develop a text-based method to measure industrial policy activity and generate the first global dataset of these economic interventions. Using natural language processing (NLP), we design an algorithm to classify industrial policy objectives—the intent to alter the composition of economic activity—directly from policy text. We apply this approach to the world's largest corpus of real-world policy documents to construct detailed measures of industrial policy at the country–sector–year level. These text-based measures not only outperform traditional proxies, such as tariffs, but also classify industrial policy activity with high accuracy (94%). Using these data, we provide the first global descriptive analysis of industrial policy activity and demonstrate why a systematic approach to measurement matters. To support further research, we provide our measures as an open-source public good here.

To illustrate the conceptual challenge for measuring industrial policy, consider tariffs, where the problem is well understood and has been widely discussed. Although sometimes used to promote infant industries (Juhász and Steinwender, 2024), tariffs are also implemented with the goal of raising government revenue (Johnson, 1951; Balassa, 1989; Cagé and Gadenne, 2018), managing terms of trade (Broda, Limao and Weinstein, 2008), or catering to political interests (Goldberg and Maggi, 1999; Gawande, Krishna and Olarreaga, 2015). Thus, a dataset of tariffs is not a dataset of industrial policies.[1] The same holds for non-tariff measures (NTMs).

1. In their assessment of the literature, Harrison and Rodríguez-Clare (2010, p. 4065) note that a fundamental problem with cross-industry studies of tariffs and growth is that there is "no evidence to suggest that intervention for IP reasons in trade even exists." Earlier work sought to address this issue by distinguishing tariffs used for fiscal revenue generation from those employed for industry promotion purposes (Lehmann and O'Rourke, 2011). Likewise, Nunn and Trefler (2010) used the skill bias of tariffs as a proxy for tariffs aimed at industry development.

Our solution to this measurement challenge is based on the observation that the *text* of policy announcements often includes information about a policy maker's goals. Tables 1 – 2 (industrial policy goals and other policy goals, respectively) provide illustrative examples from our source data, emphasizing goals (in *italics*). To see why this type of information is useful for distinguishing IP, consider three import tariff policies. One wants "to stimulate innovation and strengthen the national IT sector" (Table 1, example 1). Another wants "to increase the revenue of the government" (Table 2, example 2). A third wants to reinstate import tariffs on food staples "which were removed during the period of rising food prices" (Table 2, example 4). Although these policies use the same policy instrument, they clearly have different objectives. Of the three policies, only the first constitutes industrial policy because of its explicit goal of wanting to change the composition of economic activity towards the IT sector.

In the first part of the paper, we develop a text-based approach to measure industrial policy. Using supervised machine learning, we identify industrial policies (2010–2022) by analyzing policy text from the Global Trade Alert (GTA) database (Evenett, 2019; Evenett and Fritz, 2020), the largest inventory of commercial policies.[2]

We summarize our method in four steps. First, we define industrial policy as deliberate government action aimed at altering the composition of the domestic economy to achieve a public goal. Our definition draws from extensive historical (e.g., Johnson (1982)), economic policy (e.g., Corden (1980); Warwick (2013)), and legal (e.g.United States International Trade Commission (1983); Kapczynski and Michaels (2023)) literatures that emphasize the deliberate goals of industrial policy. Second, we manually categorize, or "label," a subset of policy descriptions in our database using a team of human annotators. We show that humans agree on the concept of industrial policy industrial policy. Third, we train two machine learning models on these labeled data—a logistic regression classifier and a large language model (LLM)—both achieving strong predictive performance. Finally, we deploy our best-performing model (LLM) to classify all policies in our database and construct text-based indicators of industrial policy activity.

In the second part of the paper, we show that our measure accurately captures industrial policy activity using various validation exercises. We first validate our text-based approach using held-out data from the supervised learning workflow. We show that standard benchmark text classifiers (logistic regression with vectorized text) outperform comparable models based on policy instruments (e.g., tariffs, subsidies,

2. Our approach follows similar efforts to measure monetary policy (Romer and Romer, 2004) and fiscal policy (Romer and Romer, 2010) shocks through qualitative assessment of policy documents. It also automates the manual data collection typically required for industrial policy case studies (e.g., Barwick, Kalouptsidi and Zahur (2019); Bai, Barwick, Cao and Li (2020); Lane (2020); Barwick, Kwon, Li and Zahur (2024a)).

Table 1: Industrial Policy Goals: Examples

	Policy Description	**Targeted Activity**	**Policy Instrument**
1	Brazil increased import tariffs for various IT and telecommunication goods *to stimulate innovation and strengthen the national IT sector.*	IT and telecommunications	Import tariff
2	The Ministry of Industry and Information Technology released a policy [...] *to boost growth in the Chinese battery industry*, particularly for automobiles.	Batteries	State Loan
3	[...] the Ministry of Information Industry (MII) of the People's Republic of China (PRC) issued a Planning Release [...] The release [...] seeks to provide guidance on *maintaining and strengthening the PRC's position in the global ship-building industry.*	Shipbuilding	State Loan
4	The Ecuadorian Executive adopted Decree 675, increasing the percentage of bioethanol in regular fuel [...] *aiming to boost biofuel consumption and production while supporting local agriculture*.	Biofuels, agriculture	Price stabilization
5	The government of Egypt increased for USD 1,000, its flight subsidies for international charter flights [...] [the] *core scope [of this program] was to boost Egyptian tourism overall.*	Tourism	Production subsidy
6	The German Federal Government published the Artificial Intelligence (AI) Strategy, *aiming to increase competitiveness and secure responsible AI development in Germany and Europe.*	AI	State aid, unspecified
7	[...] government of Japan approved [...] supplementary subsidy *to support exports of goods identified under the agricultural export expansion strategy* [...].	Selected agricultural products	Other export incentives
8	Nigeria's Federal Executive Council approved a new national automotive policy *to strengthen the automotive sector* and limit imports of used cars.	Automobiles	Import tariff
9	The South African Executive launched the Green Economy Accord [...] to *promote the development of the Green Economy.*	Green activities	State loan
10	[...] US Administration enacted the CHIPS and Science Act of 2022 *to boost American semiconductor manufacturing, research, workforce development, and advanced wireless communication technologies*.	Semiconductors	Multiple

Notes: The table shows excerpts of policy descriptions and instrument types from the Global Trade Alert database, which we use in this study. The text that refers to the goals of the policy has been italicized by us.

export loans). Augmenting these text-based classifiers with policy information does not improve predictive performance. This finding confirms the long-standing concern in the literature that policy instruments do not correspond well to industrial policy. Furthermore, we show that large language models, which incorporate the semantic content of policy text, outperform our benchmark text-based logistic classifiers, which discard textual information. Together, these results confirm that policy instruments

are weak proxies for industrial policy activity, and show richer textual models outperform conventional NLP techniques.

We next validate our text-based models across multiple dimensions. First, we test for face validity, showing that the most predictive words (e.g., "development," "promote," "technology") align with expected policy language. Second, we demonstrate that our large language model learns coherent industrial policy concepts. Using Testing with Concept Activation Vectors (TCAV), we show that the model captures industrial policy goals within key layers of the neural network. Third, we conduct out-of-sample validation with retaliatory sanctions imposed on Russia after its 2022 full-scale invasion of Ukraine. These sanctions allow us to test the model's temporal generalizability and its robustness to novel, previously unseen events.[3] Finally, our measures are validated by subsequent research: Goldberg, Juhász, Lane, Lo Forte and Thurk (2024) and Barwick et al. (2024a) report substantial overlap between our industrial policy classifications and their analyses of global semiconductor and electric vehicle policies, respectively.

In the third part of the paper, we present new findings around contemporary industrial policy practice and highlight three emerging patterns. First, industrial policy represents a significant share of policies in the GTA (44–62% of policies with identifiable goals), with prevalence rising steadily since the 2010s. Second, industrial policies, unlike other policies in our data, disproportionately target sectors where a country already holds a dominant position in export markets. No similar targeting pattern exists for non-industrial policies, indicating distinctiveness not only in declared goals (by construction) but also in implementation. Together, these results demonstrate that industrial policies are both quantitatively important and qualitatively unique.

Our third finding is that contemporary, real-world industrial policy deviates from conventional wisdom. Where early empirical work focused on tariffs, we find that subsidies and export-oriented measures are far more common. In fact, import tariffs are not among the most-used industrial policy instruments. Furthermore, the vast majority of import tariffs are not used for industrial policy reasons. Taken together, this shows that tariffs are a poor measure of industrial policy.

Similarly, in contrast to its traditional emphasis in development economics (e.g., Dervis and Page (1984)), our findings suggest that industrial policy is more widely used in high-income countries relative to developing economies. Thus, contrary to much literature evaluating individual episodes of industrial policy, the typical industrial policy today is not used by a country behind the technology frontier and is

3. Our model overwhelmingly classifies these sanctions as non-industrial policies, confirming its ability to accurately classify policy language not encountered during training.

not targeted towards an infant industry. In fact, the opposite is true: typical industrial policy today is deployed by high-income economies and targets a sector in which a country has a revealed comparative advantage. These findings highlight the added value of systematic measurement.

We contribute to three strands of literature. First, we advance the emerging empirical research on industrial policy in three ways: (1) We introduce a disciplined, systematic measurement approach that can be applied across domains. (2) We provide the first global dataset on industrial policy activity from 2010 to 2020, making our model (LLM) and data public resources for future empirical scholarship; for example, Barwick, Kwon, Li, Wang and Zahur (2024b) applies our data to cross-country studies of innovation in electric vehicles, and Goldberg et al. (2024) to learning-by-doing in semiconductors. (3) We conduct the first global, cross-country, cross-industry analysis of industrial policy activity. Our study complements recent measurement efforts using fiscal data, such as the careful accounting-based work of Criscuolo, Gonne, Kitazawa and Lalanne (2022) for select OECD nations and the comparative study by DiPippo, Mazzocco and Kennedy (2022) of seven industrial economies in 2019.

Second, our method demonstrates the potential of using text and alternative data to address broader measurement challenges in the trade policy literature. Goldberg and Pavcnik (2016) argue that measurement is a major bottleneck to understanding the impact of trade policy, especially when "actual" policy changes are not recorded. Our work joins efforts by Estefania-Flores, Furceri, Hannan, Ostry and Rose (2024), who extract trade restrictiveness from IMF reports, and Teti (2020), who apply computational methods to produce improved measurements of tariffs.

Finally, we contribute to the literature that applies natural language processing (NLP) to measuring economic phenomena (D'Orazio, Landis, Palmer and Schrodt, 2014; Gentzkow, Kelly and Taddy, 2019; Grimmer, Roberts and Stewart, 2022; Ash and Hansen, 2023). We provide a framework for combining supervised learning and large language models to construct data on economic concepts that are otherwise difficult to measure. Our paper is the first to apply these tools to outstanding questions in economic policy: industrial policy. We do so by using the tools of supervised machine learning for "concept detection," which classifies complex constructs (e.g., ideology, populism). To validate our model, we incorporate recent advances in machine learning (Kim, Wattenberg, Gilmer, Cai, Wexler, Viegas and Sayres, 2018) and demonstrate that our model captures meaningful semantic representations of economic policy. Thus, our paper provides an additional tool, beyond dictionary-based approaches to measurement used in economics, such as seminal examples by Baker, Bloom and Davis (2016) and Hassan, Hollander, Van Lent and Tahoun (2019).

The paper is structured as follows. The next section introduces our definition of industrial policy and our approach to measurement. The third section presents our source data. The fourth section describes our methodology, and the fifth validates our text-based approach and the classifiers. In section six, we use the newly constructed measures to document basic stylized facts about contemporary industrial policy and show why a systematic approach to measurement matters. The final section concludes.

2. Measuring Industrial Policy: Definition and Approach

This section introduces our general framework for measuring industrial policy. We begin by presenting our formal definition, which draws from an extensive body of policy scholarship. Next, we demonstrate how to operationalize this definition by systematically using policy text to distinguish industrial policies from other government actions. Finally, we illustrate our approach through concrete examples, emphasizing how we use policy objectives, rather than instrument type, to identify industrial policy activity.

2.*A*. Defining Industrial Policy

We define industrial policy as intentional government action aimed at altering the composition of a domestic economy to achieve a public goal. This definition builds on an extensive body of literature and captures three key dimensions of industrial policy.

i. First, industrial policy is a *political* action (Juhász and Lane, 2024) pursued by governments and implemented by states. This excludes actions taken by non-governmental actors (e.g., NGOs), firms, and other private entities. The political component of industrial policy is perhaps the most ubiquitous feature across all definitions (Lindbeck, 1981; Warwick, 2013; Criscuolo et al., 2022).

ii. Second, industrial policy is *intentional*: governments seek to alter the composition of economic activity *to achieve specific goals*. Historically, industrial policy aimed to modernize economies and drive structural change, often by promoting manufacturing (Juhász and Steinwender, 2024). More recently, goals such as facilitating the net-zero transition, enhancing supply chain resilience, and achieving strategic autonomy in key sectors have gained prominence (Juhász, Lane and Rodrik, 2024). As Chalmers Johnson explains, "[t]he very existence of industrial policy implies a strategic, or goal-oriented, approach to the economy" (Johnson, 1982, p. 19).

This goal-oriented, or intentional, aspect is the *sine qua non* of industrial policy. The deliberate "intention to alter the structure of the economy" (Warwick, 2013, p.15) is a recurring theme in definitions of industrial policy over decades. Intentionality is also central to current empirical analyses of industrial policy (Harrison and Rodríguez-Clare, 2010; Lane, 2021) and is evident across disciplines: legal studies (Kapczynski and Michaels, 2023), trade policy practitioners (United States International Trade Commission, 1983; Boonekamp, 1989), industrial economics (Ferguson and Ferguson, 1994, p.137), and public policy research (Bendick Jr. and Ledebur, 1981; Dubnick and Holt, 1985; Goldstein, 1986).

Consider some concrete examples of intentionality and goals used in definitions. "Industrial policy can be broadly defined as the *deliberate attempt by a government to influence the level and composition of a nation's industrial output*" (our emphasis) (Boonekamp, 1989, p.14). Similarly, Dubnick and Holt (1985, p.116) adopt a definition that views industrial policy as "inherently both intentional and active." Kapczynski and Michaels (2023) argue that industrial policy "involves *deliberate attempts to shape sectors of the economy to meet public aims* writ broadly" (our emphasis). Krugman and Obstfeld (1991) conceptualize industrial policy as attempts by a government to "encourage resources to move into particular sectors *that the government views as important to future economic growth*" (our emphasis). According to Chang (1994), "industrial policy is aimed at particular industries (and firms as their components) *to achieve the outcomes that are perceived by the state to be efficient for the economy as a whole*" (our emphasis). Similarly, Pitelis (2006) defines industrial policy as a set of "measures taken by a government and aiming at influencing a country's performance *towards a desired objective*" (our emphasis).

iii. Third, industrial policies exhibit *specificity*. Nearly all definitions emphasize their role in altering the structure of the economy (Warwick, 2013), and thus they favor certain activities over others (see Juhász et al. (2024)). Most famously, industrial policies may target specific industries or sectors. However, many policies cut across traditional sectoral boundaries. For example, South Korea's 1960s export-led policies broadly promoted export *activity* rather than targeting specific sectors (Lane, 2021). Similarly, recent green industrial policies promote green economic activity across sectors, such as encouraging green technologies or supporting battery supply chains.

Like intentionality, specificity has a long precedent in early discussions (Diebold, 1980; Congressional Budget Office, 1983), seminal analyses of scope (Corden, 1980; Lindbeck, 1981), and widely used definitions (Pack and Saggi, 2006) (see Warwick (2013, p.15)). Our sector-agnostic approach aligns with the independent conceptual work by Criscuolo et al. (2022) and the OECD. Practically, our broad definition means

that our dataset can be used to explore more precise studies, such as semiconductor policy (Goldberg et al., 2024) or green automotive policy (Barwick et al., 2024b).

We now turn to our approach, or how we take this definition to data.

2.B. Approach and Assumptions

Our approach uses policy text to distinguish industrial policies from policies with different objectives. As noted above, the goal of a policy has been an essential element of defining industrial policy. For our purposes, policy descriptions often include explicit language about their aims—whether industrial policy or otherwise—making classification feasible.

We take the language of policy descriptions as given; when state actions announce plans to boost specific economic activities, we classify these as having an industrial policy goal. We do so for several reasons. In explaining these rationales, we distinguish between the "primary" measures derived from text and downstream questions about their impact, veracity, and scope.

First, this paper demonstrates that industrial policy can be classified from textual sources, using a third-party dataset filtered for credibility and *de jure* policy (see Section 3). Although this dataset represents the state of the art for studies on global trade policy and non-tariff measures (NTMs), our framework can also be applied to alternative corpora. Hence, the principles of our approach are broadly applicable to other textual sources; for contemporary applications of this method, see Fang, Li and Lu (2024).

Second, despite the credibility filter for our source text, our approach does not assess bindingness or implementation success of policy. This focus is partly practical: determining whether an industrial policy is binding requires first identifying it as an industrial policy. A useful analogy can be drawn from the trade policy literature, which distinguishes between the measurement and impact of policy (Goldberg and Pavcnik, 2016). For example, studies on "tariff overhang" (Beshkar, Bond and Rho, 2015) or the material restrictiveness of NTMs (Kee and Xie, 2024) first require basic measures. Thus, we consider bindingness and related dimensions of policy as distinct research questions requiring primary inputs.

Moreover, even when implementation is imperfect or unsuccessful, state actions may still shape private actors' expectations and produce significant consequences. In fact, our method has the advantage of not selecting policies based on success along any arbitrary dimension, allowing for a more balanced assessment of a wide range of industrial policies.

Third, one may worry that even credible policies may obfuscate true, underlying policy motivations. Although political incentives, such as geopolitical concerns, can

drive obfuscation (see Kalouptsidi (2018) for discussion), there are also important reasons for signaling industrial policy goals, particularly when the policy aims to elicit private sector action, as is the case for nearly all industrial policy. Importantly, our examples below show that policymakers—despite participating in multilateral and common market agreements—are often very explicit in communicating industrial policy goals.

2.C. Examples and Application

Let us illustrate our approach to measurement. We begin with import tariffs, a domain in which measurement issues are well documented. Table 1 provides examples of industrial policies implemented through import tariffs. A Brazilian import tariff (example 1), levied on IT and telecommunication goods, aims to stimulate innovation and strengthen the national IT sector. Likewise, a Nigerian import tariff (example 8) levied on used automobiles aims to "strengthen the automotive sector." These examples demonstrate our definition of industrial policy in practice. The explicit goal of these policies is to alter the composition of economic activity in favor of specific sectors (here, IT or automobiles).

In contrast, Table 2 provides examples of import tariff policies with other, identifiable (non-industrial policy) goals. For instance, a policy from Pakistan raises import tariffs to "increase the revenue of the government," (example 2) while Ghana reintroduces tariffs on staple food items during a period of "rising food prices" (example 4). Meanwhile, Turkey eliminates import tariffs on prefabricated buildings "to meet the need for shelter caused by earthquakes" (example 6). These policies do not articulate industrial policy goals (i.e., to alter the composition of economic activity), but express other government objectives: increasing fiscal revenue, stabilizing prices for essential goods, or responding to major shocks like natural disasters.

The examples above illustrate that instruments like import tariffs can serve purposes distinct from industrial policy (Harrison and Rodríguez-Clare, 2010). Of course, some of these counterexamples are selective, targeting specific goods such as food items or building materials, and thus influence the composition of domestic economic activity. However, the key distinction is that these changes in the composition of economic activity are not the objective of the policy for non-industrial policies. Neither the selectivity of a policy nor the policy instrument itself is sufficient to identify industrial policy.

The issue above goes well beyond tariffs. Tables 1–2 demonstrate that governments use the same policy instruments to pursue both industrial policy and other objectives. Consider subsidies, loans, and financial grants, which are frequently associated with industrial policy in public discourse. Table 1 includes several examples of industrial

policies employing these instruments: Chinese state loans for shipbuilding and electric vehicle batteries, South African loans to "green the economy" (examples 2–3 and 9); an Egyptian production subsidy for tourism (example 5); German state aid for artificial intelligence development (example 6); and U.S. fiscal incentives for semiconductor manufacturing (example 10).

However, Table 2 reveals that the same instruments can serve entirely different purposes. For instance, European Investment Bank loans were provided to Portuguese firms affected by forest fires (example 7), while Vietnamese electricity subsidies aimed to mitigate COVID-19 hardships (example 9). Similarly, U.S. financial grants were allocated "to speed up the United States' recovery from the economic and health effects of the COVID-19 pandemic" (example 10), and Moroccan subsidies supported livestock producers "to alleviate the repercussions of the dry winter season" (example 5).

Having demonstrated our text-based approach, we now introduce the input data we use to classify industrial policy activity at scale.

3. Data

We apply our definition to policy text from the Global Trade Alert (GTA) project, the most comprehensive global database on state commercial policy (Evenett, 2019). GTA employs international experts and combines an automated search process with manual expert verification to identify and document new state actions. The database is designed to capture policies that *change the relative treatment of foreign versus domestic interests* (Evenett and Fritz, 2020).

The policy descriptions from GTA's inventory—examples of which are shown in Tables 1 and 2—serve as the primary input for our supervised machine learning workflow. GTA provides standardized, English-language summaries for each policy announcement, written by in-house experts. We use the April 2023 version of this continuously updated dataset.[4]

i. Inclusion and Credibility. To be included in the Global Trade Alert database, a state act must satisfy two quality criteria: it must be (i) credible and (ii) materially impactful (Evenett and Fritz, 2020). A credible act is one that has been implemented or whose future implementation is guaranteed. This excludes statements of intent, such as those made in speeches (Evenett and Fritz, 2020). While GTA refers to these

4. This version, provided by the GTA, supersedes the initial July 2020 extract used in earlier versions of the paper.

Table 2: Other (Non-Industrial Policy) Goals: Examples

	Policy Description (from GTA)	Targeted Activity	Policy Instrument
1	[...] Austrian export agency [...] opened a [...] credit line *to support companies in need of liquidity as a result of the economic destabilisation caused by the Russian invasion of Ukraine*.	All exporters	Trade finance
2	[Pakistan] Economic Coordination Committee approved several measures *to increase the revenue of the government*. An additional 1% duty has been imposed on all imported products except certain exempted items [...].	Imports	Import tariff
3	[Finland] Temporary aid scheme [...] *to support primary agricultural production in the current financial and economic crisis* ('the Temporary Framework').	Agriculture	Financial grant
4	[Ghana] [...] import duties on rice, wheat and cooking oil *which were removed during the period of rising food prices* in 2008 will be restored.	Staple food items	Import tariff
5	[Morocco] [...] Ministry of Agriculture provided 2.5 mil. quintals of barley to livestock producers at a subsidized price. [...] *to alleviate the repercussions of the dry winter season that led farmers to purchase imported grains at high prices*.	Livestock producers	Production subsidy
6	Turkey temporarily terminated the additional duties on prefabricated buildings [...] *to meet the need for shelter caused by earthquakes*.	Building materials	Import tariff
7	[...] European Investment Bank (EIB) and Banco Comercial Portugues SA signed an agreement worth EUR 75 million [...] for financing small and medium size projects [...] *impacted by forest fires in Portugal*.	SMEs	State loan
8	[...] UK government prevented Russian companies in aviation/space industry from UK-based insurance services *in response to the invasion of Ukraine by Russia.*	Russian aviation/space firms	Export ban
9	Vietnam Ministry of Industry and Trade reduced electricity price by 10%. *to ease business difficulties amid the COVID-19 pandemic.*	All	Production subsidy
10	[...] U.S. enacted American Rescue Plan Act of 2021 *to speed up recovery from COVID-19 pandemic effects*.	Various	Financial grant

Notes: Policy descriptions (excerpts) and policy instruments from the Global Trade Alert. The text that refers to the goals of the policy have been italicized by us.

entries as policy "announcements," the term should not be conflated with general political declarations or rhetoric.

Thus, the source data focus on *de jure* state action. GTA verifies measures and documents them using official statements issued by administrative institutions (Evenett and Fritz, 2020, p. 1). A typical entry is based on formal declarations by the "acting institution." In rare instances, multiple media reports are used as sources. A

meaningful policy change is defined as a state act that alters international commercial flows, whether in goods, services, investment, or labor.

ii. Scope and Coverage. GTA's coverage extends beyond the inventories maintained by multilateral institutions such as the United Nations Conference on Trade and Development (UNCTAD) and the World Trade Organization's (WTO) surveillance projects. As an independent organization, GTA avoids reliance on a country's self-reporting or compliance submissions.

To identify new policies that meet its criteria, GTA scans official government sources—including ministry websites, agency portals, and official gazettes—using automated web crawlers supplemented by expert human review. In most cases, additional leads from media outlets or industry associations are traced back to original official documentation (Evenett, 2019).

This surveillance effort captures measures beyond traditional trade policy, including both restrictive and liberalizing policies. Examples include FDI incentives, trade financing, research and development policies, and tariff reductions, demonstrating that the GTA records more than just classically protectionist measures or those deemed discriminatory under WTO rules.

The GTA is not restricted to traded commodities. Tables 1–2 illustrate the database's inclusion of policies targeting services such as tourism and insurance.

iii. Relationship to Industrial Policies. How well-suited is the GTA to capturing industrial policies? Industrial policies aim to make particular activities more attractive within an economy, often tilting the playing field in favor of domestic economic activity. As such, the Global Trade Alert captures much—perhaps most—industrial policy activity in its coverage.

Two examples illustrate the coverage of industrial policy by the GTA.

Consider consumer subsidies, such as those used to promote electric vehicle adoption (see Barwick et al. (2024a)). The GTA includes these subsidies when they incorporate local content requirements or other conditions explicitly discriminating against foreign commercial interests. The US Inflation Reduction Act exemplifies this through income tax credits for new electric vehicles meeting local content requirements.[5] In such cases, our definition of industrial policy aligns with GTA boundaries. Our definition of industrial policy excludes consumer subsidies that do not discriminate against foreign interests, as they aim to alter only consumption rather than domestic production.

As a counterexample, consider education and workforce development policies used as industrial policy instruments. The US CHIPS and Science Act uses such

5. For details, see Global Trade Alert intervention 113854.

instruments to fund workforce development and STEM education.[6] While the GTA records this act,[7] it only includes the incentives directly targeting production. These workforce development policies, though qualifying as industrial policy under our definition, fall outside GTA surveillance because they do not directly discriminate against foreign commercial interests.

iv. Quality of Coverage. The GTA seeks wide global coverage and relies on publicly available information sources (i.e., the paper trail of policies). As such, there is some concern that bigger countries, and countries with more transparent governments, have better coverage in the dataset (Evenett, 2019). In Appendix D, we compare GTA to an OECD dataset (OECD, 2024) constructed using a similar methodology, but focused on a narrower subset of policies. We conclude that despite the much broader scope of GTA, it has broadly similar coverage of policies covered by both sources. Thus, we conclude that i) the GTA is state of the art relative to what is feasible; ii) one needs to carefully examine the robustness of results to systematic mismeasurement, as any data collection effort is constrained by what policy information is available. We return to this issue in Section 6.

3.A. Units, Variables, and Summary Statistics

An observation in our data is a "measure" or "intervention." For comparability, we focus on national-level policy making. We exclude 1,371 subnational policies from the analysis.[8] We also drop data for two partial years, 2008 and 2023. We conduct our analysis on the 47,283 observations that remain after these two filtering steps. We report summary statistics in Appendix Table H.2.

Beyond the descriptions of policies, we will also use additional policy variables, or metadata provided in the source data. These variables include the announcement date; the type of intervention (e.g., a tariff, state loans, etc.); level of implementation; implementing jurisdiction; Harmonized System (HS) 6-digit code of affected sectors; and whether there was firm-level scope tied to the intervention. See Evenett and Fritz (2020) for details.

We take our definition to the data using supervised learning, which we turn to next.

6. For details, see White House Fact Sheet on the CHIPS and Science Act.
7. For details, see Global Trade Alert intervention 66540.
8. The GTA's reporting of subnational (regional) policies seems incomplete. Less than three percent of the policies in the dataset are subnational, with only 30 countries reported as having *any* subnational policies. In related work (Goldberg et al., 2024), we find that subnational, provincial policies are typically absent for Chinese semiconductor industrial policy. Given these concerns, we exclude subnational policies from the analysis. See Appendix Table H.1 for a complete distribution of the implementation levels of the policies in the database.

4. Methodology

We construct measures of industrial policy using a three-stage supervised learning process, described in this section. First, we manually label a subset of data according to our formal definition of industrial policy (Section 2.A). Second, we use this labeled data to train a classification model. Third, we apply the trained classifier to predict instances of industrial policy across our database of approximately 47,000 policies. We use these predictions to generate flow-based measures of industrial policy at the country-sector-year level. Technical documentation is provided in Appendix B.

4.A. Labeling: Annotating Subsamples

We begin by constructing training and testing data through hand-labeling a subset of policies based on our formal definition. A team of annotators codes policies using a standardized codebook that provides explicit criteria for determining whether policies meet our definition. The codebook instructs annotators to identify policy goals either through direct statements (e.g., "in order to boost domestic industry by making Egyptian cars more competitive") or implicit indicators (e.g., "China's 'Major Technical Equipment' policy grants tax-free imports to firms in certain sectors involved in the production of said equipment."). See the codebook in Appendix I for complete details.

Research assistants (RAs) from Columbia University, the University of Oxford, and the University of British Columbia hand-label 2,932 policies—approximately 6% of the dataset's 47,283 observations. These observations are randomly selected and stratified by measure type. See Appendix A for details on the annotation process.

Four RAs independently evaluate each policy, assigning one of three labels: "industrial policy," "not industrial policy," or "not enough information" (NEI). Following machine learning conventions, the NEI category captures cases where the text provides insufficient content about the target class. This proves particularly useful for sparse policy references (e.g., brief mentions of tariff-line changes), allowing our classifiers to focus more precisely on distinguishing between industrial policy goals and other policy objectives in the main corpus.

Labels are assigned through majority voting. 36% of annotated descriptions contain identifiable policy goals with industrial policies accounting for 21% of hand-labeled cases. Although policy summaries are not explicitly designed to capture policymakers' goals, such information frequently appears in the text. For the 101 ambiguous cases where annotators were evenly split, co-author Réka Juhász provided expert adjudication.

A critical step in developing our supervised learning algorithm involves establishing that human coders can consistently identify industrial policy goals according to our definition. We measure intercoder reliability using two standard metrics: Krippendorff's alpha and Conger's kappa. As detailed in Appendix A, our six rounds of annotation yielded values between 0.6 and 0.8, demonstrating both initial agreement and improved convergence over successive rounds as coders gained experience.

4.B. Training: Large Language Model and Logistic Regression Baseline

We next train classification models using our annotated data to map policy documents into one of three predicted categories (classes): industrial policy, not industrial policy, or not enough information. For our main model, we use a trained BERT (Devlin, Chang, Lee and Toutanova, 2019) (Bidirectional Encoder Representations from Transformers) large language model (LLM), which we compare against a logistic regression classifier. We use the latter as our transparent baseline benchmark. We describe each model in detail in Appendix B.

Our BERT model and baseline logistic regression classifier use fundamentally different methodologies to classify policy text. The logistic regression model employs a traditional bag-of-words approach, representing documents as sparse, high-dimensional vectors of term frequencies. Specifically, we use TF-IDF (Term Frequency–Inverse Document Frequency) features based on unigrams and bigrams (see Appendix B.2). We then train a regularized logistic regression model on the TF-IDF vectorized representations of the labeled policy text. While this approach is transparent, it disregards potentially important contextual information, such as word order and semantic meaning.

By contrast, BERT has the advantage of a pre-trained architecture. It processes documents as sequences of tokens, explicitly modeling the contextual relationships between words within the sequence. This allows BERT to capture complex semantic meanings and syntactic dependencies. Because BERT is pre-trained on large corpora of text, it acquires a general understanding of language structure and meaning before being fine-tuned for specific tasks. Training then involves fine-tuning the pre-trained model on our labeled data for the classification task.

BERT is an early example of what are now broadly referred to as large language models. Architecturally, it is an encoder-only model, unlike generative models such as GPT or Claude.[9] Instead, BERT is optimized for natural language understanding tasks, including sentiment analysis and text classification.

9. Encoders map raw inputs (e.g., tokenized text) to contextual representations; decoders autoregressively generate new tokens.

While newer LLMs and language representation models exist—including BERT variants—we use the baseline BERT (`BERT-base`) model for several reasons. First, BERT remains a strong and widely adopted benchmark for evaluating novel architectures and techniques. Second, whereas many common proprietary large language models are challenging to interpret, BERT's encoder-only architecture (Appendix Figure G.1) is comparatively well-studied and straightforward. This transparency has spawned a dedicated literature, referred to as "BERTology," which systematically investigates BERT's internal mechanisms (see Rogers, Kovaleva and Rumshisky (2020)). Thus, we adopt BERT as our baseline representation model and apply recent interpretability techniques to analyze its decision-making process (Section 5.D).

The structural differences between LLMs like BERT and logistic regression create a key trade-off. BERT's sophisticated architecture typically achieves superior predictive accuracy, particularly when prediction requires incorporating textual nuances (e.g., policy objectives), yet at the cost of reduced interpretability. Although generally less accurate, logistic regression provides transparent coefficient weights that directly indicate which terms drive classifications.

For our main BERT classifier, we employ a baseline, pre-trained `BERT-based-uncased` model (Hugging Face, 2025), which we fine-tune (train) for a three-class classification task. The computational environment, GPU, and libraries are described in Appendix B.1. As a benchmark, we use a regularized variant of logistic regression—logistic regression with L_1 (Lasso) regularization—for the same classification task. We describe the logistic pipeline, including tokenization and TF-IDF vectorization, in Appendix B.2.

For model training, we randomly partition our labeled dataset into three stratified subsets: $\mathcal{D}_{\text{train}}$ (65%) for training, $\mathcal{D}_{\text{val}}$ (20%) for validation, and $\mathcal{D}_{\text{test}}$ (15%) for held-out testing. We use $\mathcal{D}_{\text{train}}$ to estimate model parameters and $\mathcal{D}_{\text{val}}$ to tune hyperparameters for both models. The test set $\mathcal{D}_{\text{test}}$ is reserved for evaluating out-of-sample performance of both the logistic regression and BERT classifiers. To address class imbalance, we apply standard oversampling during both training and validation.

We follow best practice and select BERT and logistic models using hyperparameter tuning. For BERT, we vary learning rate, batch size, number of training epochs, and weight decay.[10] Given BERT's computational demands, we implement a Bayesian sampling algorithm to identify optimal hyperparameters, running on an NVIDIA GH200 GPU. Details of this procedure are provided in Appendix B.1. Specifically, we use a Bayesian Tree-structured Parzen Estimator (TPE) algorithm, which explores

10. The optimal hyperparameters identified for BERT model training are a learning rate of 6.0593×10^{-5}, a batch size of 32, 4 training epochs, and a weight decay of 3.0229×10^{-6}.

the hyperparameter space by learning from prior evaluations (Bergstra, Bardenet, Bengio and Kégl, 2011; Akiba, Sano, Yanase, Ohta and Koyama, 2019).[11] For logistic regression, we apply a conventional grid-search procedure with k-fold cross-validation to select the regularization strength, as described in Appendix B.2.

Following convention, we train the final BERT model and benchmark logistic classifier on the combined training and validation sets, $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$, and evaluate the final performance using the held-out test set $\mathcal{D}_{\text{test}}$. Throughout the paper, we use the F1-score as our primary evaluation metric for both training and performance assessment. The F1-score is a standard measure in machine learning, providing a single statistic for comparing models. It balances precision and recall equally and remains robust in the presence of class imbalance. See Section 5.A for further details.

4.C. Prediction: Constructing Text-Based Indices

Finally, after training, we take our best-performing model (highest global F1 accuracy) and use it to construct indicators of industrial policy language. Specifically, we take our preferred fine-tuned large-language model, BERT, to predict instances of industrial policy language throughout our entire policy dataset ($\sim 47,000$ policies). For each observation in the dataset, our model gives an indicator denoting each policy's class (industrial policy goal, no industrial policy goal, or not enough information).

Our text-based measures resemble non-tariff measures used in the trade policy literature (see Goldberg and Pavcnik (2016)). In addition to denoting new industrial policy (policy flows) activity, these indicators serve as inputs for calculating coverage ratios (e.g., Malouche, Reyes and Fouad (2013)), estimates of subsidies (e.g., Goldberg et al. (2024)), trade restrictiveness indices (Looi Kee, Nicita and Olarreaga, 2009), and *ad valorem* equivalencies (e.g., Cadot, Gourdon and van Tongeren (2018)). Before we turn to our dataset, however, we explore how our BERT model performs in predicting industrial policy activity and validate our text-based approach.

5. Model Results: Performance and Validation

This section establishes our model's predictive performance and presents validation evidence. We first show the strong predictive performance (accuracy and F1-score of 94.1%) of our preferred Large Language Model (BERT). Next, we validate our model through four complementary tests: i) We show that textual features outperform standard policy-type heuristics in predictive power. ii) We establish face validity by analyzing interpretable features from our logistic regression baseline. iii)

11. We use the Python `Optuna` optimization framework (version 4.2.1) to implement the TPE algorithm.

We apply TCAV analysis to demonstrate that BERT learns our concept of industrial policy. iv) Finally, we use the 2022 Russian invasion of Ukraine to test both temporal generalization and construct validity.

5.A. Performance: Large Language Model v. Logistic Baseline

We evaluate the performance of our best-performing Large Language Model (BERT) and logistic regression model using a sample of labeled test data, unseen by our models during training. We consider standard metrics of model performance: precision, recall, accuracy, and F1 score, the latter of which equally weights precision and recall.[12] Recall, also known as the probability of detection, measures the model's ability to correctly identify true policy instances. Precision measures the probability that an instance identified as policy is, in fact, industrial policy.

Table 3 shows that our final model reliably classifies policy on unseen, labeled test data. The BERT model achieves an average F1 score of 93.7% across classes and an overall accuracy of 94.1%, outperforming the logistic regression model (90.9% for F1 and 91.6% accuracy, respectively). Table 3 also provides a breakdown of model performance across the three policy classes. For industrial policy, our target class, the LLM, particularly outperforms logistic regression (91.3% versus 87.1% F1, respectively).

Table 3: Predictive Performance of Three-Class Models on Test Data

Model	Class/Metrics	Precision	Recall	F1 Score	Support
Large Language Model (BERT)	IP Goal	0.913	0.913	0.913	104
	No IP Goal	0.959	0.934	0.947	76
	Not Enough Information	0.947	0.954	0.950	260
	Accuracy			0.941	440
	Macro Avg	0.940	0.934	0.937	440
	Weighted Avg	0.941	0.941	0.941	440
Logistic (Benchmark)	IP Goal	0.867	0.875	0.871	104
	No IP Goal	0.971	0.882	0.924	76
	Not Enough Information	0.921	0.942	0.932	260
	Accuracy			0.916	440
	Macro Avg	0.920	0.900	0.909	440
	Weighted Avg	0.917	0.916	0.916	440

Notes: This table reports the predictive performance of our main three-class model on a held-out test sample of annotated data. We assess performance by comparing model predictions to human-coded labels using a labeled test set $\mathcal{D}_{\text{test}}$. We report results from the final BERT classifier alongside the benchmark logistic classifier (with L_1 regularization). Precision, Recall, and F1-score are reported for each class. Macro Average refers to the unweighted mean of metrics across the three classes; Weighted Averages are weighted by class size. Accuracy is calculated across all classes.

12. Formally, Recall = $\frac{\text{TP}}{\text{TP+FN}}$ and Precision = $\frac{\text{TP}}{\text{TP+FP}}$. F1 is a weighted combination of each: $F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}}$. Accuracy refers to the overall share of correct predictions: Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$.

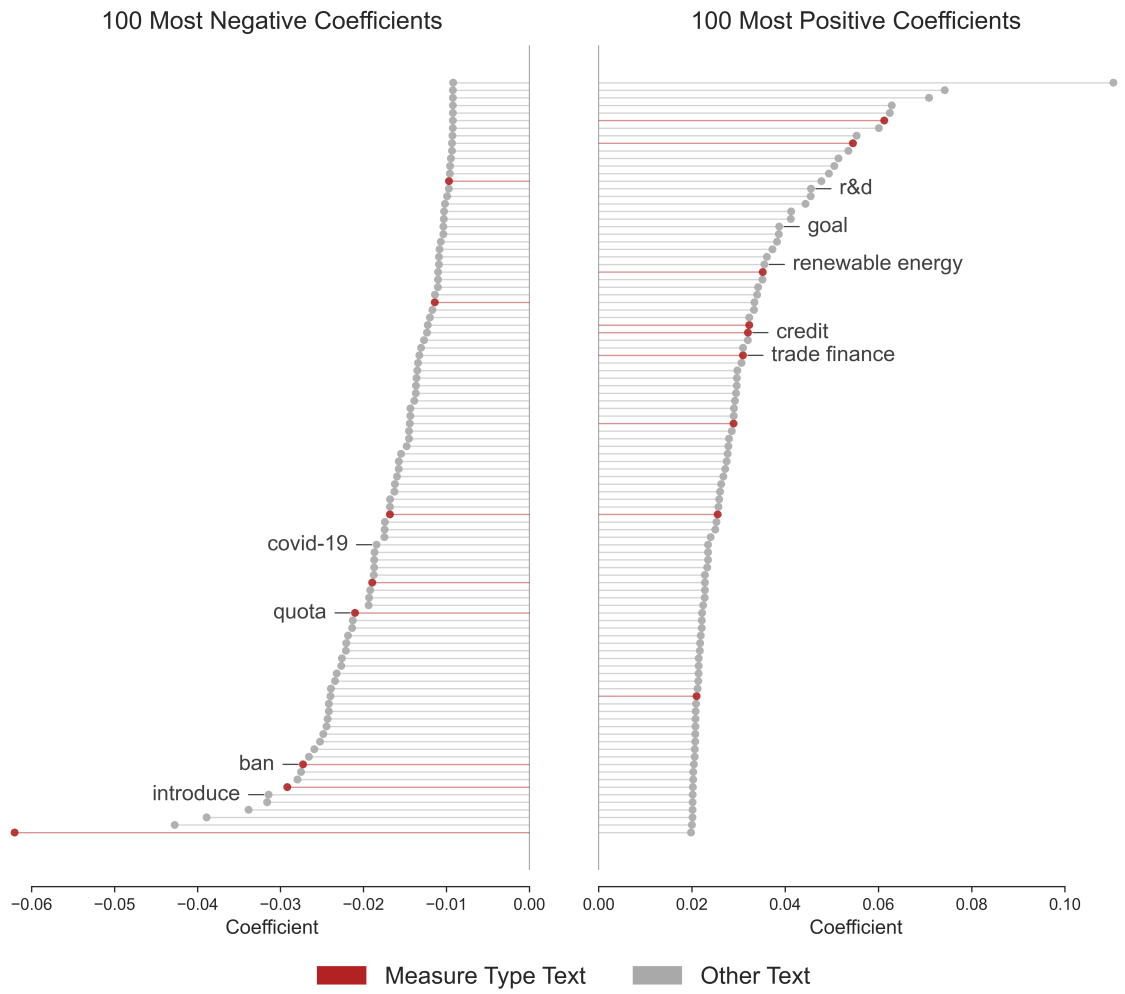


Figure 1: Most Predictive Coefficients for Industrial Policy and Coefficients Related to Measure Type

Notes: The figure displays the 200 most predictive coefficients for classifying industrial policy, consisting of the 100 most negative and 100 most positive values. Estimates are drawn from our baseline three-class logistic regression classifier (L_1 regularized). Coefficients associated with Measure Type are highlighted in red.

5.B. Validating Our Approach: The Predictive Power of Policy Text

We demonstrate that textual data are valuable for identifying whether a policy reflects industrial policy objectives, particularly when compared to common heuristics such as policy type. The predictive value of textual features is evident in the performance of unigrams and bigrams from our baseline logistic regression model, as illustrated in Figure 1, which plots the 200 most predictive coefficients. The coefficients highlighted in red correspond to unigrams or bigrams explicitly linked to measure type categories (e.g., credit, trade finance), as defined by the UN's policy classification system (MAST, or Multi-Agency Support Team, codes).

Measure type alone is insufficient to reliably identify policies with industrial policy objectives. Figure 1 illustrates that while information about policy measure type contributes to classification, a broad array of textual features is also important, even in the simplest classifier.[13] Appendix Figure G.2 further shows that policies classified as industrial policy and those classified as having other goals fall within the same MAST categories. This suggests that categorical policy variables, by themselves, may be inadequate for predicting industrial policy content accurately.

Table 4 formally evaluates the predictive power of textual information relative to heuristic indicators such as policy categories. It compares the performance of textual features and categorical policy variables using our baseline logistic classifiers. Panel (a) of Table 4 reports results from three logistic regression models predicting our target class—industrial policy—on held-out test data.

The first model uses only measure type as a predictor. We employ the MAST measure categories from the GTA dataset, applying one-hot encoding to convert categorical variables into numerical values. This measure type-only model is tuned using k-fold cross-validation and regularized with L_2 (ridge) to retain all policy features. We compare this model (measure type only) to the baseline text-only classifier (from Table 3), and to a third model (text and measure type) which augments the text model with policy category. Both text-based models use L_1 regularization.

The results in Table 4 show that policy category information alone performs poorly relative to textual features. The measure type-only classifier achieves an F1-score of 71.4%, while the text-only model reaches an F1-score of 87.1%. A nonparametric McNemar test comparing prediction differences on the test set confirms that the performance gap is highly significant ($p < .00001$; $\chi^2 = 34.2$). Moreover, adding policy type features to the text model does not improve predictive accuracy; the combined model performs slightly worse (86.3% for F1), though this decline is not statistically significant ($p < .48$; $\chi^2 = 0.5$). These findings suggest that even relatively simple text-based classifiers outperform policy category indicators in identifying industrial policy activity.

5.C. Validation: Face-Validity and Logistic Classifier

We next assess the face validity of our text-based classification approach, first using our logistic regression model. While BERT and logistic regression have fundamentally different architectures, we first use the logistic model as an interpretable benchmark to examine which textual features are most predictive of industrial policy. This

13. Note that the ranking of individual tokens is different from their contribution, in practice, to actual predictions. Individual tokens alone typically exert limited influence; predictions result from the combined effect of many terms.

Table 4: Predictive Performance of Text Versus Measure Type

(a) Predictive Performance for Industrial Policy Across Models

Class	Logistic Model	Precision	Recall	F1-score
Industrial Policy	Measure Type Only	0.747	0.683	0.714
	Text Only	0.867	0.875	0.871
	Text and Measure Type	0.850	0.875	0.863

(b) Testing For Model Differences (McNemar Test)

Comparison	Statistic (Chi-Square)	p-value
Text vs. Measure Type	34.22	0.0000
Text vs. Text and Measure Type	0.50	0.4795

Notes: Panel (a) reports the performance of three logistic regression models for the target class, industrial policy. Each optimized model is evaluated using Precision, Recall, and F1-score on the held-out test set. The comparison includes: (i) the text only logistic classifier (baseline), (ii) the measure type-only classifier, and (iii) a combined model that incorporates both textual features and measure type. All text-based models are regularized with L_1, while the measure type-only model is regularized with L_2. Panel (b) presents results from McNemar tests, which evaluate differences in predictions between the text only model (i) and the models with measure type (ii and iii).

interpretable baseline helps triangulate the behavior of our more complex large language model, which we validate in Section 5.D.

Figure 2 presents the top twenty unigrams and twenty bigrams with the highest estimated coefficients for the target class, industrial policy, as identified by the logistic classifier. These terms serve as the strongest predictors that a given text will pertain to industrial policy. As recommended in the literature, examining the largest coefficients offers insight into model behavior, while smaller coefficients should be treated with caution (Gentzkow et al., 2019). Notably, the most influential terms include intuitively relevant language such as "development," "technology," "exporter," "boost," "promote," "growth," "increase competitiveness" and "long term." These results suggest that the model captures semantically meaningful content closely aligned with industrial policy objectives.

5.D. Validation: BERT Learns Industrial Policy Objectives

We demonstrate that our fine-tuned BERT model meaningfully incorporates economic concepts—"industrial policy goals"—into its classification decisions using Concept Activation Vectors (CAVs) (Kim et al., 2018). While logistic regression allowed direct evaluation of feature importance (Section 5.C), neural networks like BERT

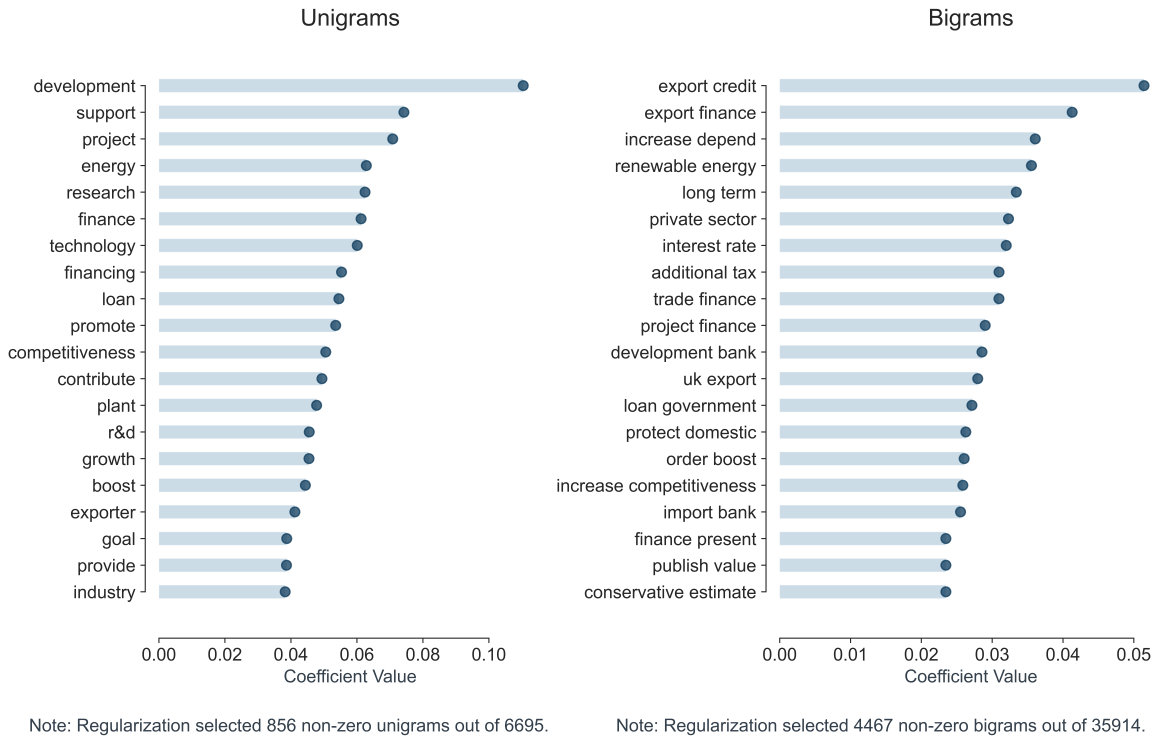


Figure 2: Top Predictive Coefficients (Unigrams and Bigrams) of Industrial Policy for Logistic Classifier

Notes: This figure displays the top 20 coefficients most predictive of industrial policy from our baseline logistic regression classifier. Separate panels show the top 20 unigrams and top 20 bigrams. The coefficients are estimated using L_1-regularized logistic regression, with the regularization strength selected via grid search and k-fold cross-validation. Each plot reports the number of features retained by the L_1 penalty.

operate as "black boxes" where internal decision processes remain opaque. Testing with Concept Activation Vectors (TCAV) addresses this challenge (Castelvecchi, 2016) by quantifying how abstract, human-interpretable concepts influence model predictions. Our TCAV analysis reveals that the model internalizes domain knowledge in deeper network layers where abstract representations are formed. Crucially, we validate that this concept learning is semantic rather than superficial by testing robustness across different data distributions and pooling methods.

1. Methodology. TCAV quantifies how much a high-level concept influences a model's classification decisions by examining how the model internally represents inputs. We first learn a direction in the model's activation space that corresponds to a concept ("industrial policy goals"), then measure how sensitive the model's output is to changes in that direction—that is, whether moving in the concept direction increases the model's confidence in a particular prediction. Full details are in Appendix C.

A neural network processes input text $\mathbf{x}$ through a series of layers. At each layer ℓ, it creates an activation vector $f_\ell(\mathbf{x})$, a high-dimensional representation of the text.

BERT consists of 12 transformer layers, where earlier layers capture syntax while later ones capture abstract meaning.

A concept C ("industrial policy goals") is defined as a curated set of example texts, X_C. These concept texts produce specific activation patterns, which we compare against activations from random baseline texts (negative examples, X_N). We train a linear classifier to separate these two sets of activations $f_\ell(\mathbf{x})$. The normal vector to the classifier's decision boundary is our Concept Activation Vector (CAV), $\mathbf{v}_C^\ell$—the direction in activation space pointing toward concept C.

The CAV measures a concept's influence on predictions, or conceptual sensitivity, for a specific class k and input $\mathbf{x}$:

$$S_{C,k,\ell}(\mathbf{x}) = \nabla h_{k,\ell}\big(f_\ell(\mathbf{x})\big) \cdot \mathbf{v}_C^\ell$$

where $h_{k,\ell}$ maps activations from layer ℓ to a prediction score (logit) for class k, and the gradient $\nabla h_{k,\ell}$ indicates the direction that most increases the likelihood of predicting class k. The dot product measures alignment between this prediction-enhancing direction and our concept direction. When $S_{C,k,\ell}(\mathbf{x}) > 0$, moving the activation toward the concept increases the likelihood of classifying as k.

The TCAV score quantifies a concept's overall importance as the fraction of examples from our dataset, $\mathcal{X}_k$, where the concept positively influenced the prediction:

$$\text{TCAV}_{C,k,\ell} = \frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} \mathbb{I}\big[S_{C,k,\ell}(x) > 0\big], \tag{1}$$

Scores range from 0 to 1, where 0.5 indicates random chance and scores above 0.7 suggest the concept systematically contributes to classifying inputs as class k. In practice, we robustly calculate TCAV using multiple randomly generated negative sets paired with concept C.

2. Implementation. We implement the TCAV analysis in four steps, calculating the TCAV scores for the deeper layers of BERT. We provide details of the implementation in Appendix C.

i. Generate Concept Examples. The concept set is designed to be internally coherent, semantically consistent, and distinguishable from unrelated concepts. We generated 300 examples for our target concept, C: industrial policy goals (e.g., "to promote the growth of the manufacturing sector" and "boosting the horticulture export market"). Concept set examples were generated using large language models (Claude 3.7 Sonnet and ChatGPT 4o), seeded with prototype examples. To ensure conceptual breadth, we deliberately spanned multiple sectors (services, agriculture, manufacturing) and

activities (R&D and innovation). Appendix Table H.3 presents 20 representative examples, with details in Appendix C.1.

To estimate the concept activation vector, we curated diverse "negative examples" that are clearly distinguishable from the target concept. These negative sets serve as a baseline against which the unique activations associated with industrial policy goals can be identified. We constructed sixteen negative sets, $\mathcal{N}_1, \ldots, \mathcal{N}_{16}$, each containing 300 strings of formal, declarative English sampled from our source corpus while filtering out industrial policy vocabulary. Although this in-distribution sampling controls for stylistic similarity, it risks conceptual leakage where negative examples are unintentionally contaminated with industrial policy semantics. We therefore also constructed sixteen out-of-distribution negative sets sampled from open-source corpora (Wikipedia, arXiv, PubMed), detailed in Appendix C.1.

ii. Extract Activations from Deep Layers. Next, we pass the concept examples C and negative sets ($\mathcal{N} = \bigcup_{i=1}^{16} \mathcal{N}_i$) through our fine-tuned BERT model to extract activation vectors $f_\ell(\mathbf{x})$. Specifically, we process examples through the upper layers 11 and 12, which are known to capture abstract semantic relationships prior to classification (Tenney, Das and Pavlick, 2019).[14] Appendix C.2 provides further details on the extraction of activations from deep layers.

iii. Train CAVs on Activations. Third, for each layer ℓ, we use the activations from Step ii) to train an L_2-regularized logistic regression classifier that distinguishes between concept and non-concept activations. The normalized vector of coefficients from this classifier defines the Concept Activation Vector, $\mathbf{v}_C^\ell$, which corresponds to the normal vector of the decision boundary separating the two groups.

iv. Compute TCAV Scores. Finally, we use the learned CAVs from Step iii) to assess their influence on real examples from the prediction dataset. For each class k (industrial policy, not industrial policy, not enough information), we sample 300 high-confidence predictions $\mathbf{x} \in \mathcal{X}_k$. To ensure high quality, we randomly draw examples from the top 90th percentile of the model's class-specific logits. For each such example, we compute its conceptual sensitivity $S_{C,k,\ell}(\mathbf{x})$ and aggregate these scores at the class–layer level to produce the final measure, $\text{TCAV}_{C,k,\ell}$. We repeat this calculation for each of the sixteen concept–negative pairs and report the mean aggregated score for each layer.

3. Results: Concepts Impact Predictions. Our analysis demonstrates that the fine-tuned BERT model systematically associates the concept of "industrial policy goals" with the correct class. Figure 3 reports average TCAV scores for layers 11 and 12, aggregated across 16 negative sets. Each panel shows the mean TCAV for all

14. Our index differs slightly from machine learning conventions that index from 0.

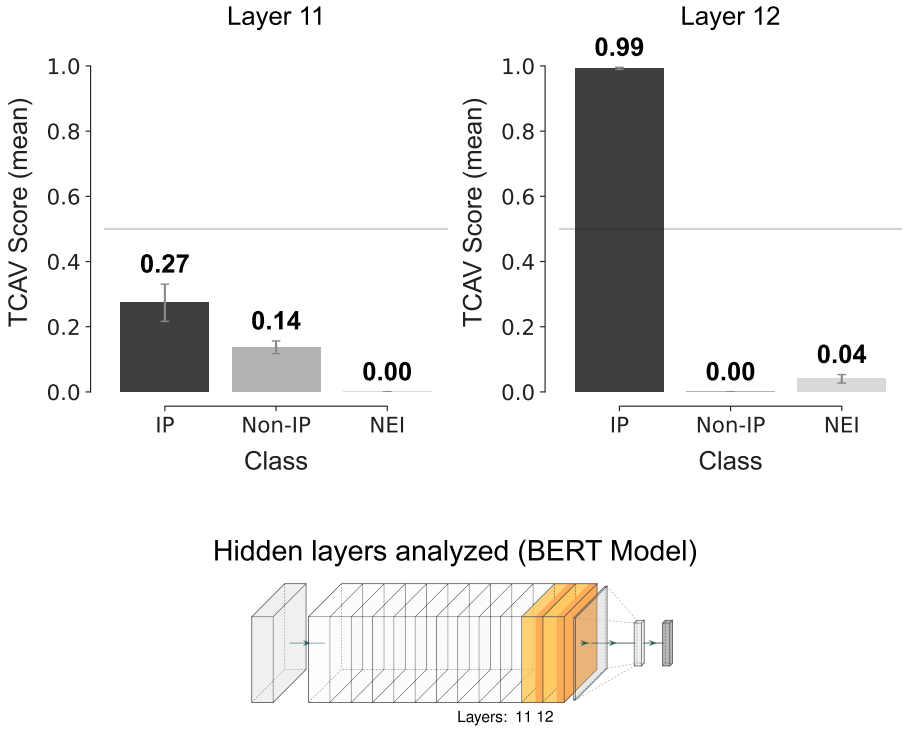


Figure 3: Average TCAV Scores for 'Industrial Policy Goals' for Deep Layers, by Class

Notes: This figure presents TCAV results for our fine-tuned BERT model, evaluating the importance of the concept "industrial policy goals" in classifying policy text. The analysis focuses on BERT's final encoder layers (11–12), which capture high-level semantic abstractions prior to classification. Each panel corresponds to one encoder layer. Bars represent TCAV scores for each class: industrial policy, not industrial policy, and not enough information (NEI). Scores were computed by comparing the concept C against 16 distinct negative concept sets ($\mathcal{N}_1, \ldots, \mathcal{N}_{16}$). The y-axis reports the average TCAV score, and error bars indicate the standard deviation across these 16 comparisons. High TCAV scores for the industrial policy class suggest that the model consistently internalizes this concept as a key feature for classification. In contrast, the concept exhibits low predictive relevance for the non-industrial policy classes.

three classes: 'industrial policy' (dark gray), 'not industrial policy,' and 'not enough information' (light gray). The line at 0.5 denotes scores that are as good as random. Specifically, for the target 'industrial policy' class, the TCAV score in Figure 3 is nearly perfect (0.99) at the final consequential transformer layer (12). In contrast, the TCAV scores for the 'not industrial policy' and 'not enough information' classes remain at approximately zero.

These findings indicate that BERT's final layer directly encodes the abstract concept as a decisive feature for classification. Layer 12 absorbs most task-specific information for classification. Since BERT's prediction head directly consumes layer 12 activations (see Appendix Figure G.4), this alignment suggests the concept directly influences classification decisions.

4. Robustness: Validating Genuine Concept Learning. To ensure our results reflect genuine concept representation rather than token-specific artifacts, we test robustness to different 'pooling' methods—how the model's activations are processed for computing the CAV. We describe this in Appendix C.2. Our calculations use the [CLS] token, BERT's special summary token, which is also directly used for classification.[15] However, our results are identical when using 'mean pooling,' aggregating information across tokens. The TCAV score is ≈ 1 for layer 12 in Appendix Figure G.3. This consistency between the summary token and mean pooling results suggests that "industrial policy goals" are broadly distributed across text, rather than concentrated in individual tokens.

To validate genuine concept learning rather than spurious pattern matching, we test our results with alternative negative sets drawn from out-of-distribution data. If the model were relying on surface-level features of our source data, TCAVs may change dramatically when using negatives with different stylistic characteristics. Appendix Figure G.4 shows near-perfect TCAV scores at layer 12 (both for mean pooling and CLS). The fact that scores for layer 12 are identical suggests that the model has learned to distinguish "industrial policy goals" based on genuine semantic content rather than dataset artifacts.

5.E. Validation: Temporal Generalization and Russian Sanctions as an Out-of-Sample Experiment

We next evaluate the model's ability to classify entirely new policy events, using out-of-sample events not present in our training data. We use Russia's invasion of Ukraine in February 2022 and the ensuing international sanctions as a validation test: i) whether our model generalizes to novel, unseen events (temporal generalization), and ii) whether it correctly distinguishes industrial policy from other interventions based on underlying objectives versus surface features (construct validity).

The post-invasion sanctions episode provides a useful experiment. While sanctions significantly affect industries, trade, and factor flows—features common to industrial policies—they fundamentally differ in their objectives, targeting geopolitical pressure rather than shaping the composition of domestic production. Thus, sanctions offer ground truth for testing whether our model has learned the conceptual boundaries of industrial policy.

Our codebook was developed in September 2021, prior to the invasion, and all annotated policies used in model training were drawn from a random subset of

15. BERT attaches a "special" [CLS] token to tokenized text, which acts as a summary of textual input. See Appendix C.2.

policies implemented through 2020. The model therefore had no exposure to or foreknowledge of the sanctions regime that would emerge.

Figure 4 plots the classification of sanctions policies targeting Russia from 2018 through 2023. The dashed line marks the February 2022 invasion, after which sanctions (grey) spike dramatically. Notably, our model classifies 95% of these sanctions as either having non-industrial policy objectives or lacking sufficient information for classification. Thus, our model avoids a false positive trap. This suggests the model has successfully learned to identify industrial policy based on its objectives rather than merely flagging targeted market interventions.

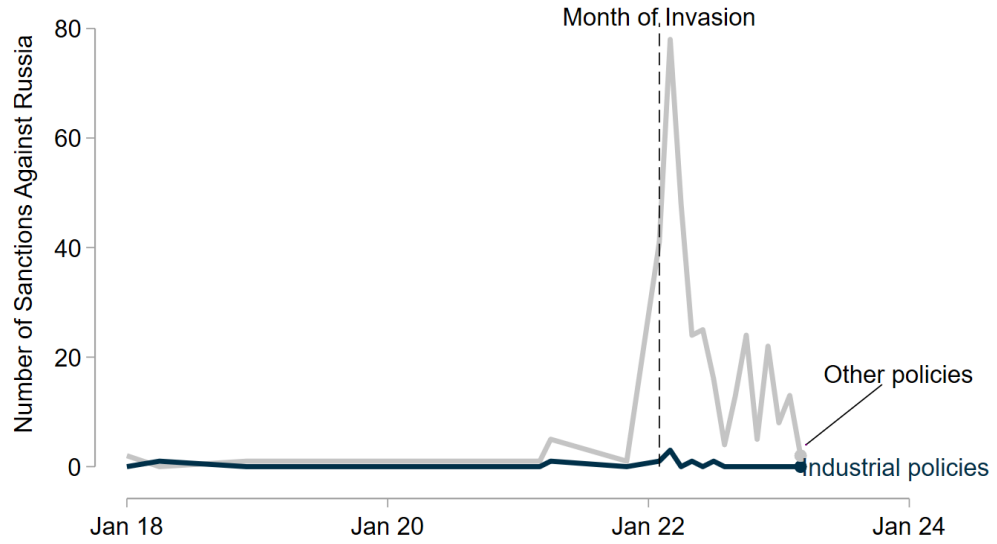


Figure 4: Russian Sanctions Labeled as Industrial Policy and Not Labeled as Industrial Policy

Notes: The figure reports how our model classifies policies related to sanctions in the context of Russia's invasion of Ukraine. We identify likely sanctions on Russia using two criteria: (1) the affected countries include Russia, and (2) the policy description contains one or more of the following terms: "sanction," "invasion," "frozen," "aggression," or "illegal." As of April 2023, 374 such policies have been published in the GTA. Of these, 95% are classified by the model as either pursuing other policy goals or lacking sufficient information (NEI class).

6. Stylized Facts

In this section, we use our measure of industrial policy activity to examine basic patterns in real-world policy implementation. Although the theoretical literature on industrial policy offers numerous predictions about both *what* industrial policy

should target and *how* it should be conducted, there remains little systematic empirical evidence on these questions. We structure our empirical analysis around four guiding questions: 1. How much industrial policy activity is there? 2. How is industrial policy deployed? 3. Do some types of economies use industrial policy more intensively than others? 4. What types of sectors are most frequently targeted?

We present our empirical results using multiple strategies to address variations in the way industrial policy activity is reported across countries and implementing agencies. The GTA reports policies at varying levels of granularity: when firm-level data are available, each firm's support is recorded as a separate policy; when such detail is unavailable, a single aggregate policy is reported (see Appendix E for examples). Additionally, there is concern that GTA may double-count some policies by recording them at different implementation stages. For example, in countries with granular reporting, a policy may be counted once at the time of announcement and again when support is disbursed to specific firms.

We show: i) simple total counts; ii) "national" policy counts (excluding policies implemented at the firm level); and iii) counts of the "implementing agency" using data on the institutions deploying industrial policies, from Juhász and Lane (2024); Field (2024). The first, baseline, measure (i) shows all industrial policy activity. This measure is difficult to compare across countries if policies are reported at different levels of granularity. The second (ii) and third (iii) measures address these issues. The second measure (ii) excludes any industrial policy reported as being implemented at the firm level (as provided in the GTA source data).[16] This approach mitigates concerns about double-counting and inconsistent reporting standards across countries or programs by completely discarding information reported at the the firm level. For this reason, we consider measure (ii) a particularly conservative approach that likely excludes many legitimate policies we wish to include.

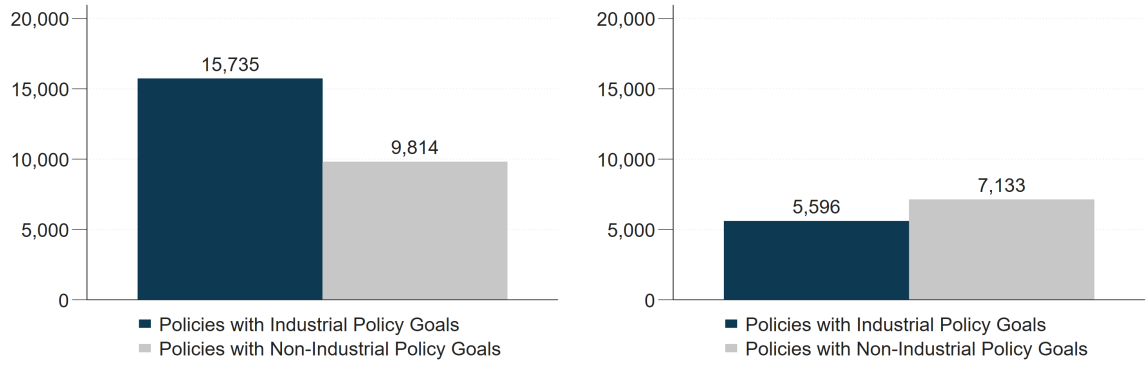
The third measure (iii) addresses the same concerns by using the implementing agency as the unit of analysis (e.g., the ministry of finance providing capital injections, a publicly owned financial institution providing export credit, or a national rail corporation offering subsidized rail freight for targeted domestic sectors). Specifically, we use the policy-implementing agency-year as our unit of analysis, recording policies implemented by an agency in the same year as a single policy. This method accommodates varying reporting standards across policy programs and countries without entirely discarding individual policies. For example, an export loan program implemented by a public financial institution and disbursed to many firms in the

16. We use the term "national" policy to refer to policies that do not only target a small set of firms. This is different from the implementation level of policies reported in Appendix Table H.1. Recall that subnational policies are excluded from the data.

same year is counted once, regardless of whether individual loans are separately enumerated in the data.

We construct the third measure (iii) by extracting implementing agencies from the policy description text using implementing agency name recognition techniques (Juhász and Lane, 2024; Field, 2024). Specifically, we use OpenAI's ChatGPT API to identify the name of the implementing agency, followed by manual cleaning to harmonize the names of implementing agencies and distinguish public from private agencies. This procedure is detailed in Appendix F and Appendix Table H.4 contains examples. We extract the implementing agencies only for policies classified as industrial policy.[17]

Each of the three measurement approaches has distinct strengths and limitations. We present results using all three methods. We do this both to assess robustness and to illustrate how different data constructions influence empirical patterns. The raw data may be affected by variation in national reporting practices. The second approach likely omits too much information, potentially under-representing actual policy activity. The third approach—based on the number of implementing agencies—may reflect differences in state administrative capacity rather than policy intensity. For these reasons, comparing across approaches provides a more comprehensive view of practice.



(a) Total Policies (b) National Policies

Figure 5: Model Classification of Policies with Identifiable Goals

Notes: This figure shows the results of the classification from our BERT model. This exercise excludes policies that did not have identifiable goals based on the preferred BERT model.

17. We limit this extraction to industrial policy because identifying implementing agencies is substantially more complex for other types of policies, which may lack a lead agency or omit this information from the description.

6.A. Fact 1 - Industrial policy is important and on the rise

The first finding is that industrial policy is a quantitatively important policy tool. We gauge this by determining the fraction of commercial policies in the GTA dataset that have industrial policy goals. Figure 5 shows that 44-62% of policies with identifiable goals are industrial policy.[18]

Industrial policy is also on the rise. In Figure 6, all evidence points to an expansion of industrial policy activity over the past decade. Panel (a) shows that the raw count of industrial policies has increased more than thirtyfold between 2010 and 2022. National-level industrial policy activity (Panel b) increased fifteenfold, indicating that the GTA is not merely capturing policies at a more granular level over time. Based on (Panel c), in 2022, the number of public agencies announcing at least one new industrial policy worldwide was close to an order of magnitude higher than in 2010.[19]

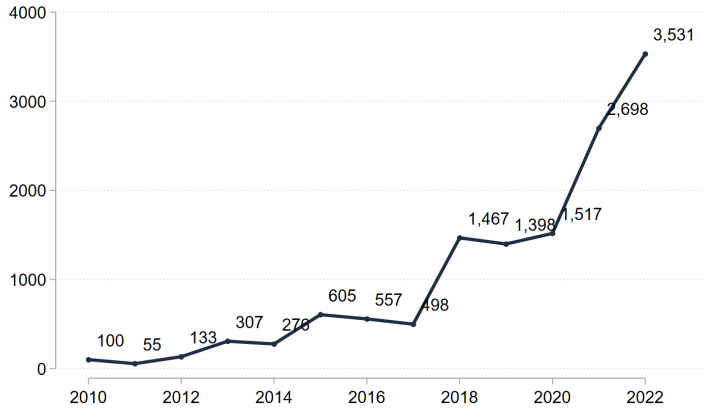
In the appendix, we show further evidence that it is unlikely that our results are driven by the GTA's increased ability to collect data, as the share of industrial policy among all policies has also increased (Appendix Figure G.5). These results suggest that industrial policy is a relatively common government intervention. We also find evidence supporting the widely held view that industrial policy has been on the rise in the 2010s (e.g., Stiglitz, Lin and Monga (2013); Cherif and Hasanov (2019)).

6.B. Fact 2 - Contemporary industrial policy favors export promotion and subsidies, not import tariffs.

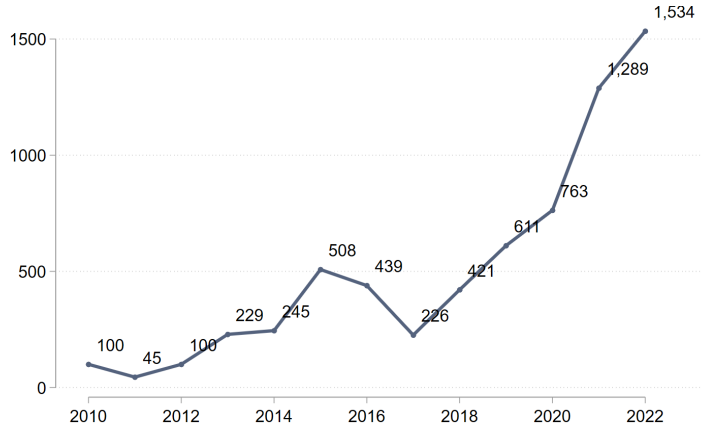
The second finding relates to the policy instruments used to deploy contemporary industrial policy. Figure 7 shows that subsidies and export-related measures (e.g., trade financing) dominate contemporary industrial policy activity, regardless of how it is measured. Even based on our most conservative estimates, IP activity deployed via subsidies and export-related measures is nearly an order of magnitude more common than import tariffs. Panels (b) (national policies) and (c) (implementing agencies) account for the concern that our count-based measure of IP activity may overstate the importance of measures, such as subsidies, relative to tariffs, as the former may be reported on a firm-by-firm basis in some countries.

Moreover, subsidies and export-related measures are the most common instruments of industrial policy across the income distribution. Appendix Figure G.6 reports the top instruments of IP, splitting countries into groups by income quintile.

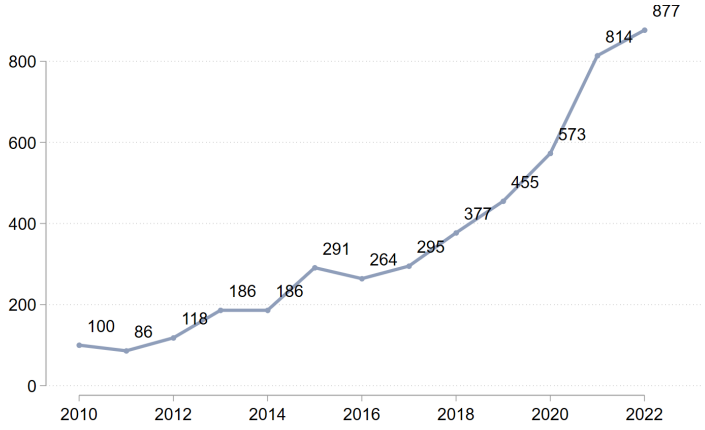
18. For this exercise, we exclude policies whose goals could not be identified (i.e., the "not enough information" group). Approximately 47% (national policies) - 54% (all policies) of the policies in the GTA had identifiable goals based on our preferred BERT model.
19. To assess time trends, we count unique agencies implementing at least one industrial policy in each year.



(a) All Industrial Policies



(b) National Industrial Policies



(c) Agencies Implementing Industrial Policies

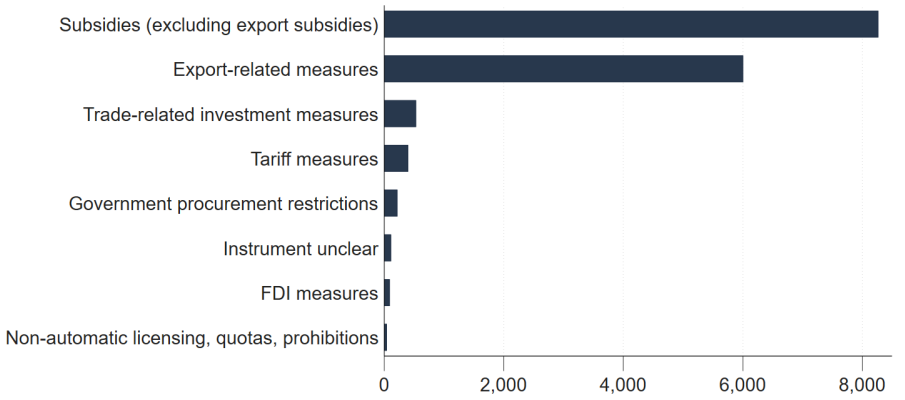
Figure 6: Time Trend of Industrial Policy Activity

Notes: This figure shows the trend of industrial policy from 2010 to 2022. We follow GTA guidance and use only policies recorded by the GTA in the same year that they were announced for this exercise. This is due to the substantial backfilling of data that is a living dataset. By using only policies recorded in the same years as they were announced, we ensure the comparability of data across both more distant and recent years. The figure is presented as an index with the year 2010 set as the base year (indexed to 100). All subsequent values reflect changes relative to this baseline.

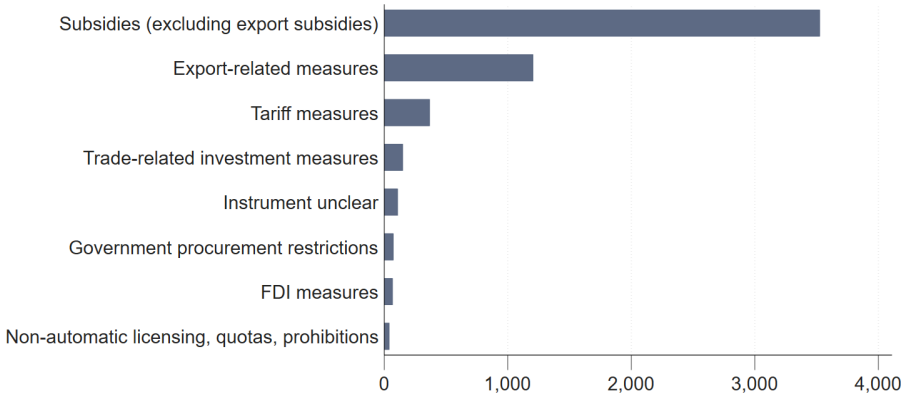
We split all countries in our data into income quintiles based on 2010 GDP per capita data from the World Bank. These two instruments are the most common, regardless of how we measure IP activity. While these instruments dominate in high-income countries, low- and middle-income countries also use import tariffs, FDI measures, and trade-related investment measures (e.g, local content requirements) relatively more. This broader mix of policy instruments, particularly the relatively higher share of import tariffs, would be consistent with the more constrained fiscal capacity faced by lower-income countries. The relatively larger use of FDI measures may also signal that in lower-income countries, IP may be used to attract investment from the technology frontier.

To understand the specific types of instruments used for IP, we report finer categories in Appendix Figure G.7. This figure reveals that most export-related IP measures are deployed via trade-financing, and, to a lesser extent, financial assistance in foreign markets. Tax-based export incentives and export subsidies (which are, in general, banned by the World Trade Organization) are much less common. Put differently, export-related IP measures tend to operate by providing financing rather than directly incentivizing exports. In the case of subsidies, industrial policy is most often deployed via state loans, financial grants, and loan guarantees. Production subsidies, capital injection and equity stakes, and tax or social insurance relief are less common. Similar to export-related measures, subsidies are thus also typically deployed by providing or supporting financing for firms.

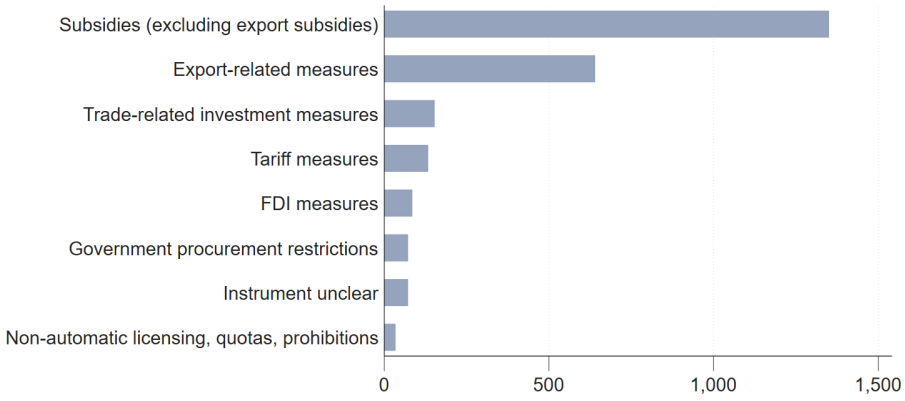
These findings challenge long-held assertions about industrial policy. First, import tariffs give a misleading and incomplete picture of industrial policy. They mislead because most tariffs do not seem to be used for industrial policy goals (Appendix Figure G.2). Moreover, they are incomplete because the majority of industrial policies are not deployed via tariffs (Figure 7). Second, industrial policy and protectionism are often conflated. While tariffs may have dominated industrial policy in earlier periods, today's typical industrial policy is not protectionist. Instead, much of it aims to boost export competitiveness, a point we explore further below. Finally, these patterns highlight that modern industrial policy requires fiscal resources and high administrative capacity. Specifically, states need sufficient fiscal revenue to subsidize firms and promote exports, plus the administrative capacity to identify which firms to support. These dimensions of state capacity provide context for interpreting our next stylized fact.



(a) All Policies



(b) National Policies



(c) Implementing Agency-Policy Instrument Pairs

Figure 7: The Instruments of Industrial Policy

Notes: The charts show the top eight most-used policy instruments by all the measures. Below eight, IP activity is too small to display clearly. The most-used policy instruments by each measure of IP activity are the same, and the excluded policy instruments are the same for each measure of IP activity, too. The excluded MAST chapter codes are: Price-control measures, Migration measures, Capital control measures, Finance measures, Intellectual property, Contingent trade-protective measures. For Panel (c), Implementing Agency–Policy Instrument Pairs, we calculate the number of agencies implementing at least one industrial policy via each policy instrument in each year, and then sum these counts across years.

6.C. Fact 3 - Industrial policy is heavily used by high-income economies

Our third finding is that, although industrial policy is common, its use varies across countries. This variability is seen in the raw data. Appendix Figure G.8 plots the distribution of industrial policy by income quintile, and shows a strong positive correlation between measures of industrial policy activity and income. The data show that countries in the top income quintile deploy six to eighteen times as many industrial policies as countries in the lowest income quintile.

To systematically examine whether industrial policy use is correlated with income, we regress a country's (log) total number of industrial policies on a set of binary indicator variables that denote the income quintile of each country in our dataset. More formally, we estimate cross-sectional regressions of the form:

$$\log(1 + \text{IP}_c) = \alpha + \sum_{g=2}^{5} \beta_g \mathbb{1}_{\{c \in g\}} + \gamma' X_c + \epsilon_i \tag{2}$$

where c indexes country, g indexes income quintile $g = 2, ..5$, and X_c are country-level control variables. The coefficients of interest are β_i, which measure the difference in IP activity relative to the excluded (lowest) income quintile ($g = 1$).

Figure 8 plots the estimated β_g coefficients for each quintile. The pattern is consistent: regardless of how we measure IP activity, higher income quintiles are associated with greater use of industrial policy. The coefficient of interest is large, and the difference is statistically significant for the fourth and fifth income quintiles—representing high-income countries. The baseline estimates (with no controls) suggest that total IP activity in the fourth and fifth income quintile is 500-2000% greater than in the poorest income quintile. In robustness checks, we control for the size of the country, measured as the log of population, to account for the fact that larger countries may have more policies (we use 2010 population data from the World Bank). Additionally, the (log) count of exported products (at the HS6 level), controls for the diversification of the economy (data on the number of HS6 sector codes traded by each country comes from UN COMTRADE).

One explanation for these results is that we may be systematically undercounting industrial policy in lower-income countries. We investigate several ways in which this type of measurement error could enter our industrial policy dataset.

First, the GTA may track policies less accurately in low-income countries, where limited administrative capacity and government transparency create measurement challenges (See Section 3). We evaluate the scope of this type of bias by comparing the GTA to a benchmark dataset: the OECD's "Inventory of Export Restrictions on Industrial Raw Materials" (OECD, 2024). This third-party inventory tracks export

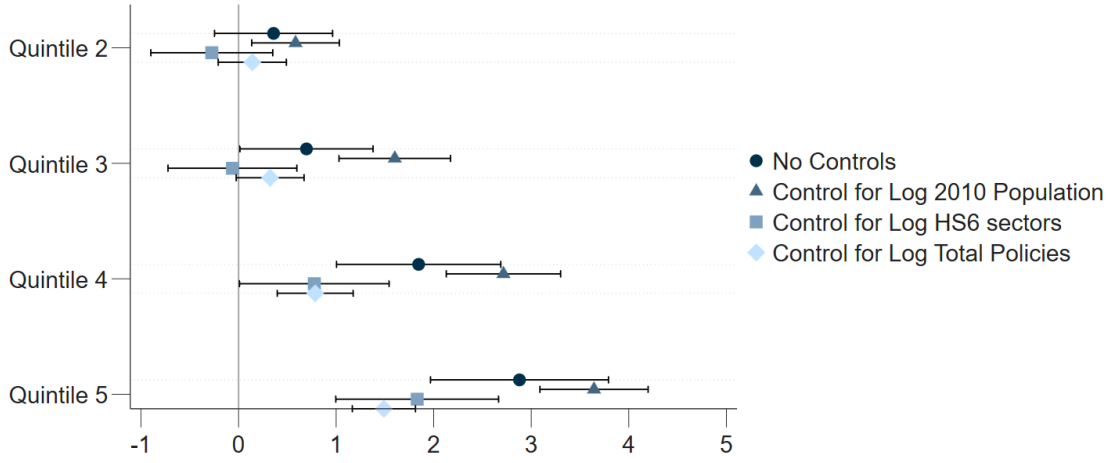
restrictions worldwide. We assess the GTA's policy coverage by hand-matching policies to the OECD's 2022 dataset for a stratified random subset of countries (see Appendix D).[20] Appendix Figure G.9 plots the share of OECD policies which are also identified in our input data: we find no evidence that lower-income countries are systematically under-reported relative to the OECD benchmark (Appendix Figure G.10 shows the correlation between match rate and income level). We note that this exercise cannot account for the fact that the paper trails of policies may be different in low-income countries, which would affect both datasets. We deal with this type of measurement challenge next.

Second, Figure 8 shows that the difference between industrial policy use at the top and bottom of the income distribution persists even after controlling for the total number of policies, although the point estimates decrease. If the GTA underreports policies in low-income economies—where paper trails are harder to find for example, this bias alone cannot fully explain the correlation between industrial policy and income. If measurement error drives our results, it must be the case that the GTA undercounts industrial policy in lower-income countries to a larger extent than other policies.

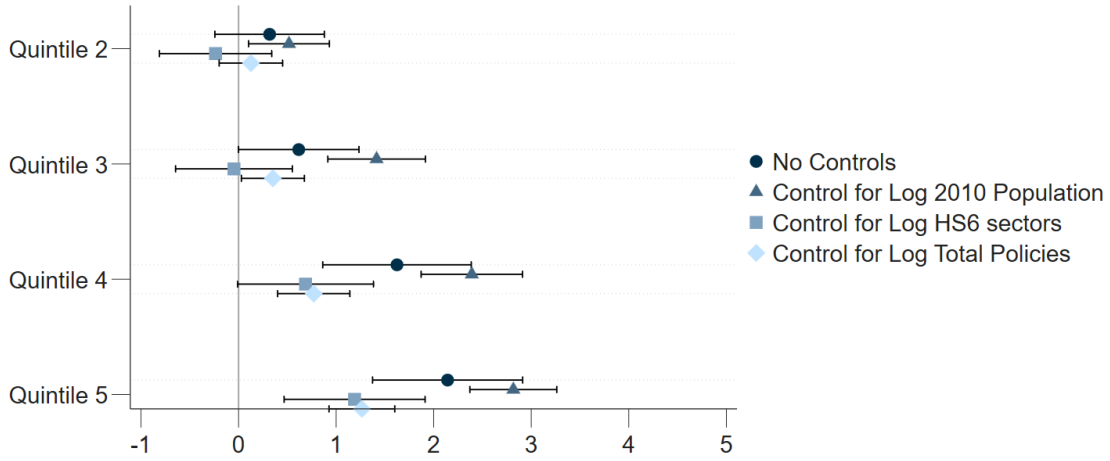
Third, the regression results also suggest that different reporting standards are unlikely to account for the patterns, as they hold across all three different measures of IP activity: Panels (a)–(c), in Figure 8.

Fourth, our data captures industrial policy flows versus stocks. This could bias our understanding of industrial policy practice across countries if, for example, low-income countries have a higher stock of policies but amend or introduce new policies less frequently. We use the same OECD dataset (on export restrictions of raw materials) to evaluate the potential scope of this bias, because the OECD data reports both stocks and flows. Appendix Figure G.11 shows that the average annual flow of policies is stable at around 20% relative to the stock across the income distribution.

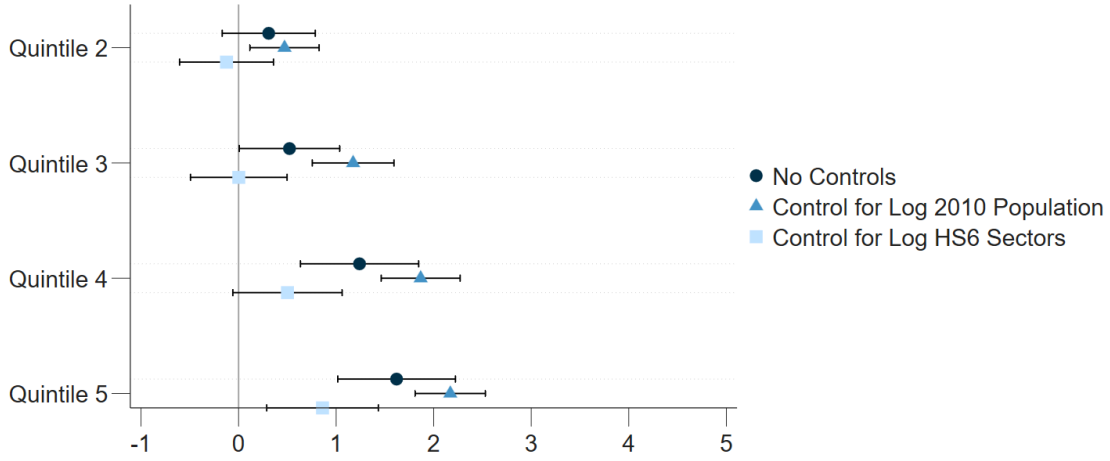
20. The OECD data focuses on a subset of policies that disproportionately affect low-income countries, making this comparison particularly well-suited for gauging potential under-reporting in these nations.



(a) All Industrial Policies



(b) National Industrial Policies



(c) Agencies Implementing Industrial Policies

Figure 8: Regression of Industrial Policy Activity on Income Quintiles

Notes: This figure plots the estimated β_g coefficients from equation 2 for each income quintile and their corresponding 95% confidence intervals. The specifications are as follows: "No controls" is the baseline estimate with no control variables, "Control for Log 2010 Population" adds a control for the log of a country's population (measured in 2010), "Control for log HS6 sectors" controls for the count of traded HS6 sectors, and "Control for Log Total Policies" controls for the log total number of policies enumerated by the GTA for the given country. Appendix Table H.5 contains the corresponding regression table.

Fifth, if our model fails to generalize—for instance, by over-specializing to policy language used by high-income countries—we may under-record industrial policies in low-income countries. We show that our baseline model, trained on diverse policies from all income levels, generalizes well across the income distribution. Appendix Table H.6 compares our "Main Model" against a High-Income (HI) Model, which is trained exclusively on data from high-income economies. The HI-Model misses 22% of industrial policy cases in low-income economies (recall = 0.78). In contrast, our Main Model achieves an F1 of 0.9 on the same income group. Moreover, the HI Model underperforms on its own turf, with an F1 score of 0.85 on high-income data compared to our Main Model's 0.92 (see Appendix Table H.6). For the remaining categories ("No IP Goal" and "Not Enough Information"), the differences between the models are less pronounced, suggesting that these classifications are less dependent on income levels. Our results are robust to restricting the test data to a common sample of policies from low- and middle-income countries, allowing for a fair comparison of model performance (see Appendix Table H.7). Together, these results indicate that our full model successfully learns patterns associated with industrial policies across the income distribution, performing well in both high- and low-income settings.

Finally, policy text from lower-income countries' policy text may simply contain less information about their goals and thus are more often classified as "not enough information."[21] Indeed, Appendix Figure G.12 shows evidence consistent with the lower information content of policy in low-income countries. For low- and middle-income countries (quintiles 1–3), over 60% of policies are classified as "not enough information", while the share of not enough information policies is as low as 20% for the highest income quintile. To understand whether we are missing industrial policy in lower-income countries due to higher rates of missing information content in these economies, we conduct a bounding exercise in which we reclassify all the not-enough-information content policies in low- and middle-income countries which *might* be industrial policy as industrial policy. Appendix Figure G.13 shows that the highest income quintile continues to have more industrial policies than the poorest countries.

In summary, the evidence presented in this section points to the fact that higher-income countries are the heaviest users of industrial policies. More precisely, high-income countries account for a disproportionate share of the *type* of industrial policy our approach is well-suited to capturing. This finding is consistent with the evidence from the previous fact, which suggests that *all* countries typically deploy fiscally and administratively intensive industrial policy. If contemporary industrial

21. Note that low-information content policies are distinct from the issue discussed above, which is about different ways of describing industrial policy goals.

policy is disproportionately deployed via fiscally and administratively costly policies everywhere, it is perhaps unsurprising that advanced economies are the ones that can afford these policies.

However, we caution that a more tailored approach to measurement in lower-income countries may find a more important role for other, less financialized instruments.[22] This caveat aside, this result provides robust evidence that high-income economies play an outsized role in actively shaping the composition of economic activity with financialized instruments.

6.D. Fact 4 - Industrial policy is correlated with revealed comparative advantage in high-income economies

Our fourth fact examines the types of industries targeted by new industrial policy activity. We explore whether industrial policy systematically targets sectors that are more or less established in international markets. Theories of industrial policy differ on which industries governments should target, but to date there is no empirical evidence on what targeting looks like in practice.

To study patterns of targeting, we merge measures of industrial policy activity with trade flow data using the United Nations Commodity Trade Statistics Database (UN COMTRADE). Trade values are reported in USD, and we consider trade flows at the Harmonized System (HS) aggregate 2-digit and 6-digit levels. We use these data to construct revealed comparative advantage (RCA) (Balassa, 1965), which is a widely used metric for measuring export specialization. It is defined as $\text{RCA}_{kc} := \left(\frac{X_{kc}}{\sum_c X_{kc}}\right) \Big/ \left(\frac{\sum_k X_{kc}}{\sum_{c,k} X_{kc}}\right)$, where X_{kc} denotes country c's exports in industry k. When a country has a revealed comparative advantage in sector k that is greater than one, it means the country is more specialized in exporting that sector than other countries on average.

We run linear probability model (LPM) regressions of the form

$$\text{IP}_{kct} = \alpha + \beta \text{RCA}_{kct} + \gamma_{ct} + \epsilon_{kct},$$

where IP_{kct} is a binary indicator variable that takes the value of one if HS sector k in country c in year t has at least one new industrial policy announcement, RCA_{kct} is revealed comparative advantage, and ϵ_{kct} is the error term. All specifications include country-by-year effects γ_{ct}. We estimate the specification at the HS2 level for the years 2010-2022, using a sample of 175 countries reported in COMTRADE. Standard errors are clustered at the country level.

22. A good example of an industrial policy we would be unlikely to capture is the "Productivity Roundtables" discussed in Juhász et al. (2024), which explicitly eschewed subsidies and deployed industrial policy through government coordination with the private sector.

Table 5 shows that sectors with higher RCA are more likely to receive new industrial policy interventions. On average, a sector with an RCA above 1 has a 1.97 percentage point higher probability of receiving a new industrial policy intervention based on the estimates from column 1. This is both a statistically significant and economically meaningful effect: on average 4.3% of sector-country pairs receive a new industrial policy intervention in any given year. These results are qualitatively similar when using the continuous (log) RCA measure (column 3). Although this is a correlation, the fact that we capture the *flow* of industrial policy aids interpretation. Thus, reverse causality, namely that sectors have higher RCA because of the new industrial policy announcement, is unlikely.

Table 5: Regression of Industrial Policy Activity on RCA

	(1)	(2)	(3)	(4)
Independent Variables	IP = 1	IP = 1	IP = 1	IP = 1
RCA > 1	0.01967***	0.00590**		
	(0.00380)	(0.00241)		
GDPpc > Median × RCA > 1		0.02692***		
		(0.00695)		
ln(RCA)			0.00194***	0.00073**
			(0.00036)	(0.00030)
GDPpc > Median × ln(RCA)				0.00259***
				(0.00075)
Observations	199968	199968	180176	180176
R-squared	0.330	0.330	0.327	0.327
Mean	0.043	0.043	0.047	0.047
# of Countries	175	175	175	175

Notes: This table shows the relationship between industrial policy targeting and revealed comparative advantage. We regress an indicator of industrial policy at the country-year-HS2 level on revealed comparative advantage (RCA). IP takes the value of 1 if, for a given country-year pair, that sector (2-digit HS) received at least one industrial policy. RCA measures are created with trade data from COMTRADE. Standard errors are clustered by country. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels. All columns include country-by-year fixed effects. Mean refers to the mean value of the dependent variable.

The results in Table 5 suggest potentially strong heterogeneity across the income distribution. The effect is much stronger for higher-income countries, shown by the interaction of RCA with an indicator variable for countries with above median income (columns 2 and 4). To further explore this heterogeneity, we split the baseline sample into different income quintiles and estimate the following specification:

$$\text{IP}_{kct} = \alpha + \sum_{i=2}^{5} \beta_i \cdot \mathbb{1}\{\text{RCA}_{kct} = i\} + \delta_{ct} + \eta_{kct}, \tag{3}$$

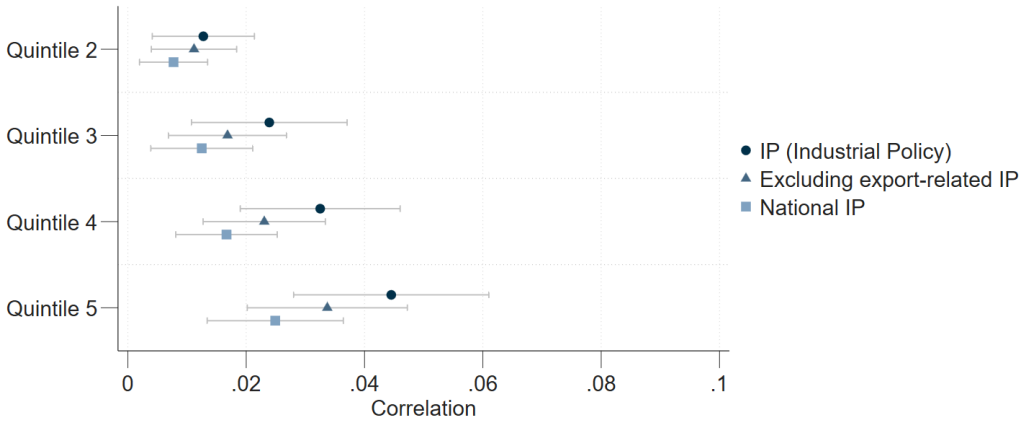
Where IP_{kct} is a binary indicator variable that takes the value of one if HS sector k in country c in year t has at least one new industrial policy announcement, β_i is the coefficient on an indicator variable that takes the value of one if RCA_{kct} is in quintile i of country c's distribution of revealed comparative advantage in year t, δ_{ct} are country-year fixed effects, and η_{kct} is the error term. The omitted category is the lowest RCA quintile, which implies that β_i captures the difference in probability of industrial policy for RCA quintile i relative to the lowest quintile of the RCA distribution.

We run this specification separately for i) the two lowest income quintiles, ii) the middle income quintile, and iii) the two highest income quintiles. Figure 9 plots the coefficients of interest.

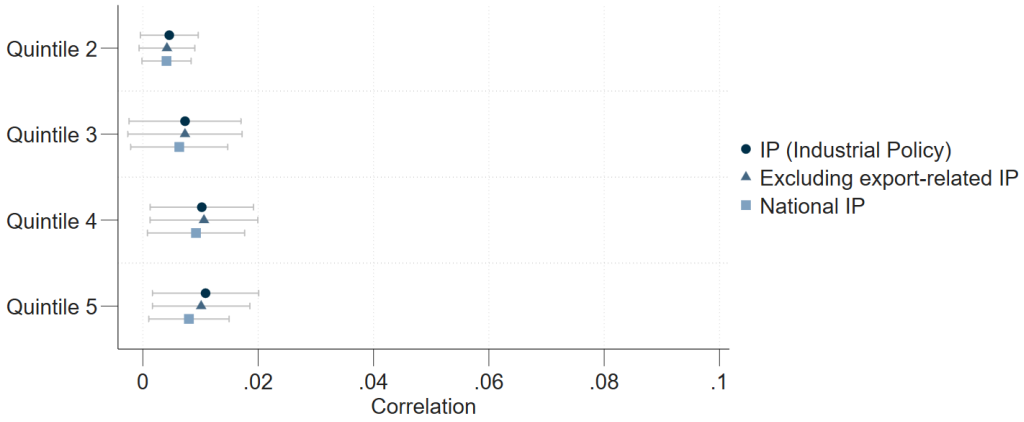
For high-income countries (Panel a), we see a strong increasing monotonic relationship between a sector's position in the RCA distribution and the probability of receiving a new industrial policy. In high-income countries, the best-performing sectors (quintile 5) are about 4 percentage points more likely to receive a new industrial policy than the worst-performing sectors (quintile 1). The results are robust to using only national policies, and to dropping all export-related policies. This latter result shows that industrial policies that do not use export-oriented policy instruments also disproportionately target a country's best-performing export sectors.

For high-income countries, we show that these results also hold when using more disaggregated (HS6) product data, reported in Appendix Figure G.14. We find an increasing, monotonic relationship even in specifications with country–year–HS2 fixed effects, meaning that within broad sectors, rich countries target their best-performing products with industrial policy.

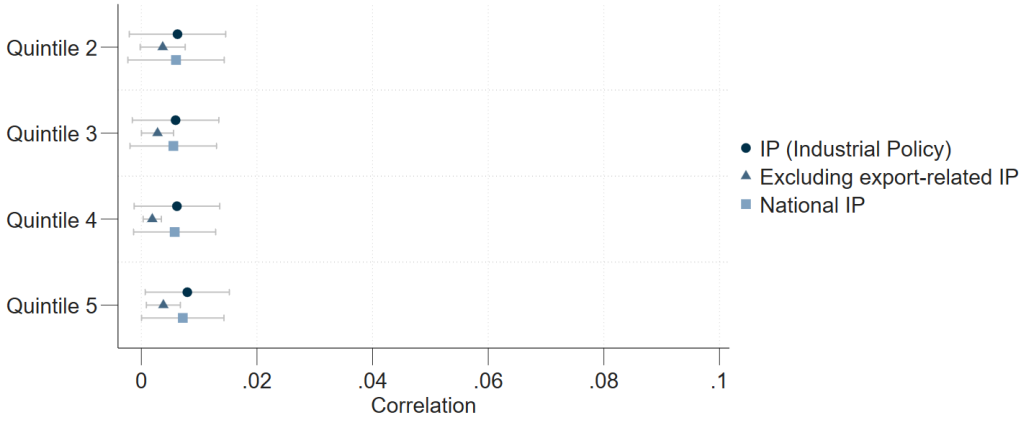
Panels (b) and (c) of Figure 9 show there is no similar effect for low- and middle-income countries. In middle-income countries, there is some evidence that better-performing sectors receive more new industrial policy, but the effect is much smaller in magnitude and does not display the same monotonic pattern as in high-income countries. For the lowest-income countries, the effect is even weaker.



(a) High-Income Countries



(b) Middle-Income Countries



(c) Low-Income Countries

Figure 9: Sectoral Industrial Policy Activity by Income Levels and Revealed Comparative Advantage

*Notes:*This figure plots the estimated β_i coefficients from equation 3 and their corresponding 95% confidence intervals. All regressions include country-year fixed effects and cluster standard errors by country. The dependent variables are as follows: "IP (Industrial Policy)" takes the value of one if a country-HS2 sector-year receives at least one industrial policy; "Excluding export-related IP" excludes industrial policies deployed via export-related measures; "National IP" takes the value of one if a country-HS2 sector-year receives at least one national industrial policy. The omitted category is the lowest quintile of the RCA distribution. High-income refers to quintiles 4 and 5. Middle-income refers to quintile 3. Low-income refers to quintiles 1 and 2.

Is the pattern of targeting found for high-income countries unique to industrial policy, or is it a more general feature of policymaking in rich countries? To answer this question, we estimate equation 3 with non-industrial policies (i.e., policies with other objectives) as the dependent variable. Figure 10 shows that policies with other identifiable (non-industrial policy) goals do *not* display the same pattern of targeting. In fact, there seems to be no relationship at all between RCA and policy for non-industrial policies. This result underscores the benefits of a systematic approach to measurement by showing that industrial policy is different from other types of policy (which may be implemented using identical policy instruments).

Although a complete exploration of these results is beyond the scope of this paper, a few remarks are in order. First, while we do not make causal claims about the pattern of targeting found in this paper, the results are more consistent with some theories of industrial policy than others. In particular, theories of infant industry predict that industrial policies should promote sectors that do not (yet) have a comparative advantage. We find no evidence in any country group for the starkest empirical prediction of this argument, which would suggest a *negative* correlation between RCA and targeting if industrial policy were trying to *defy* comparative advantage.

Other theories of industrial policy imply that policy should target sectors in which a country has already shown export viability (e.g., Hausmann and Rodrik (2003); Lin and Chang (2009)). For rich countries, the evidence is most consistent with this pattern, although it is interesting that even within broad HS2-digit sectors, countries disproportionately target their highest-performing HS6-digit products. This could be the case if *maintaining* comparative advantage at the cutting edge of technologies benefits from consistent industrial policy support. For example, research and development-intensive sectors such as advanced semiconductor manufacturing have been shown to receive ongoing industrial policy support by countries at the technology frontier (OECD, 2019; Goldberg et al., 2024).

Of course, the practice of industrial policy need not conform to any economic theory in which policy targets market failures. Equally, targeting may be driven by the political incentives of policymakers (e.g., (Juhász and Lane, 2024)). Better understanding the role of large, politically influential "superstar" exporting firms that can single-handedly shape a country's revealed comparative advantage (e.g., Freund and Pierola (2015)) and might play a role in shaping its industrial policy could be another fruitful direction for future work.

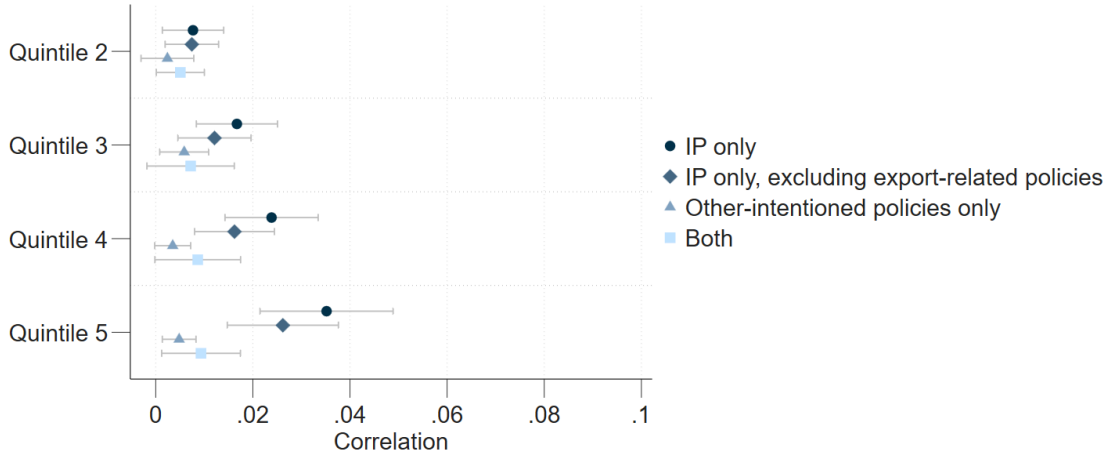


Figure 10: Regression of IP and Non-IP on Revealed Comparative Advantage in High-Income Countries

Notes: This figure plots the estimated β_i coefficients from equation 3 and their corresponding 95% confidence intervals. We estimate the regression on the sample of high-income countries only (quintiles 4 and 5 of the income distribution). The legend refers to the dependent variable used in each specification. The dependent variables are as follows: "IP only" takes the value of one when a country-HS2 sector-year receives at least one industrial policy and no other-intentioned policy; "IP only excluding export-related policies" takes the value of one when a country-HS2 sector-year receives at least one industrial policy deployed by policy instruments that are not export-related and no other-intentioned policies; "Other-intentioned policies only" takes the value of one when a country-HS2 sector-year receives at least one other-intentioned policy and no industrial policy; "Both" takes the value of one when a country-HS2 sector-year receives at least one industrial policy *and* at least one other-intentioned policy. The independent variable takes a value of one if RCA_{kct} is in quintile i of country c's distribution of revealed comparative advantage in year t. The omitted category is the lowest quintile of the RCA distribution. All regressions include country and year fixed effects and cluster standard errors by country.

7. Conclusion

In this paper, we introduce a new approach to measuring industrial policy, addressing a longstanding measurement challenge in the empirical study of state intervention. By analyzing policy language rather than relying solely on policy instruments, our text-based approach distinguishes industrial policies from other policy objectives.

Our model has strong predictive performance, achieving 94.1% accuracy and 93.7% F1-score in classifying policies. Our validation tests show that the model captures genuine economic meaning rather than spurious correlations: textual features substantially outperform conventional policy-type classifications, and interpretable features from our baseline model align with economic intuition. Furthermore, we show evidence that BERT has learned substantive policy concepts in its deeper layers.

Our data and results confront conventional wisdom about contemporary industrial policy. First, industrial policy is quantitatively significant, constituting a large share of commercial policies in our dataset. Second, contrary to traditional development economics perspectives, industrial policy is predominantly used by high-income countries, rather than developing economies. Third, industrial policies disproportionately target sectors where countries already possess a comparative advantage, not infant industries. This pattern is driven, in particular, by high-income economies. Finally, subsidies and export-oriented measures are among the most common industrial policy instruments today.

Our approach and results underscore the importance of disciplined measurement in understanding the role of state intervention in the economy. This approach demonstrates the potential of new empirical research on industrial policy across countries and sectors, allowing scholars to move beyond case studies and limited datasets. By providing open-source industrial policy measures, we support a growing research agenda on when and how government intervention shapes economic outcomes. As policymakers increasingly deploy industrial policy worldwide, systematic measurement provides the foundation for evidence-based evaluation of these economic interventions.

References

Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," in "Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining" 2019, pp. 2623–2631.

Artstein, Ron, "Inter-Annotator Agreement," in Nancy Ide and James Pustejovsky, eds., *Handbook of Linguistic Annotation*, Dordrecht: Springer, 2017, pp. 297–313.

Ash, Elliott and Stephen Hansen, "Text Algorithms in Economics," *Annual Review of Economics*, 2023, *15*, 659–688.

Bai, Jie, Panle Jia Barwick, Shengmao Cao, and Shanjun Li, "Quid Pro Quo, Knowledge Spillover, and Industrial Quality Upgrading: Evidence from the Chinese Auto Industry," Technical Report, National Bureau of Economic Research 2020.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, November 2016, *131* (4), 1593–1636.

Balassa, Bela, "Trade Liberalisation and "Revealed" Comparative Advantage," *The Manchester School*, May 1965, *33* (2), 99–123.

_ , "Tariff Policy and Taxation in Developing Countries," Technical Report, World Bank, Washington, DC 1989.

Barwick, Panle Jia, Hyuk soo Kwon, Shanjun Li, and Nahim Zahur, "Drive Down the Cost: Learning by Doing and Government Policies in the Global EV Battery Industry," Technical Report, National Bureau of Economic Research 2024.

_ , Hyuk-Soo Kwon, Shanjun Li, Yucheng Wang, and Nahim B Zahur, "Industrial Policies and Innovation: Evidence from the Global Automobile Industry," Working Paper 33138, National Bureau of Economic Research November 2024.

_ , Myrto Kalouptsidi, and Nahim Bin Zahur, "China's Industrial Policy: An Empirical Evaluation," Technical Report, National Bureau of Economic Research 2019.

_ , _ , and Nahim Bin Zahur, "Industrial Policy: Lessons from Shipbuilding," *Journal of Economic Perspectives*, 2024, *38* (4), 55–80.

Bendick Jr., Marc and Larry C Ledebur, "National Industrial Policy and Economically Distressed Communities," *Policy Studies Journal*, 1981, *10* (2), 220–235.

Bergstra, James, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl, "Algorithms for Hyper-Parameter Optimization," in "Advances in Neural Information Processing Systems," Vol. 24 2011.

Beshkar, Mostafa, Eric W Bond, and Youngwoo Rho, "Tariff Binding and Overhang: Theory and Evidence," *Journal of International Economics*, 2015, *97* (1), 1–13.

Boonekamp, Clemens, "Industrial Policies of Industrial Countries," *Finance & Development*, March 1989, *26*, 14–17.

Broda, Christian, Nuno Limao, and David E Weinstein, "Optimal Tariffs and Market Power: The Evidence," *American Economic Review*, 2008, *98* (5), 2032–2065.

Cadot, Olivier, Julien Gourdon, and Frank van Tongeren, "Estimating Ad Valorem Equivalents of Non-Tariff Measures: Combining Price-Based and Quantity-Based Approaches," Technical Report 215, OECD, Paris 2018.

Cagé, Julia and Lucie Gadenne, "Tax Revenues and the Fiscal Cost of Trade Liberalization, 1792–2006," *Explorations in Economic History*, 2018, *70*, 1–24.

Castelvecchi, Davide, "Can we open the black box of AI?," *Nature News*, 2016, *538* (7623), 20.

Chang, Ha-Joon, *The Political Economy of Industrial Policy*, London: Palgrave Macmillan, 1994.

_ , *Kicking Away the Ladder: Development Strategy in Historical Perspective*, London: Anthem Press, 2002.

Cherif, Reda and Fuad Hasanov, "The Return of the Policy That Shall Not be Named: Principles of Industrial Policy," Technical Report 2019/074, International Monetary Fund, Washington, DC 2019.

Congressional Budget Office, "The Industrial Policy Debate," Technical Report, Congressional Budget Office, Washington, DC 1983.

Corden, Warner M., "Relationships Between Macro-economic and Industrial Policies," *World Economy*, 1980, *3* (2), 167–184.

Criscuolo, Chiara, Nicolas Gonne, Kohei Kitazawa, and Guy Lalanne, "An Industrial Policy Framework for OECD Countries: Old Debates, New Perspectives," Technical Report, OECD, Paris 2022.

Dervis, Kemal and John M Page, "Industrial policy in developing countries," *Journal of Comparative Economics*, 1984, *8* (4), 436–451.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in "Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)" 2019, pp. 4171–4186.

Diebold, William, *Industrial Policy as an International Issue*, New York, NY: McGraw-Hill, 1980.

DiPippo, Gerard, Ilaria Mazzocco, and Scott Kennedy, "Red Ink: Estimating Chinese Industrial Policy Spending in Comparative Perspective," Technical Report, Center for Strategic and Internatianl Studies, Washington, DC 2022.

D'Orazio, Vito, Steven T Landis, Glenn Palmer, and Philip Schrodt, "Separating the Wheat from the Chaff: Applications of Automated Document Classification Using Support Vector Machines," *Political Analysis*, 2014, *22* (2), 224–242.

Dubnick, Mel and Lynne Holt, "Industrial Policy and the States," *Publius*, 1985, *15* (1), 113–129.

Estefania-Flores, Julia, Davide Furceri, Swarnali A Hannan, Jonathan D Ostry, and Andrew K Rose, "A Measurement of Aggregate Trade Restrictions and Their Economic Effects," *The World Bank Economic Review*, October 2024.

Evenett, Simon J, "Protectionism, State Discrimination, and International Business since the Onset of the Global Financial Crisis," *Journal of International Business Policy*, 2019, *2* (1), 9–36.

__ and Johannes Fritz, "The Global Trade Alert Database Handbook," Technical Report, Global Trade Alert 2020.

Fang, Hanming, Ming Li, and Guangli Lu, "Decoding China's Industrial Policies," 2024.

Ferguson, Paul R. and Glenys Ferguson, *Industrial Economics: Issues and Perspectives*, 2nd ed., New York, NY: New York University Press, 1994.

Field, Lottie, "The Political Economy of Industrial Development Organisations: Are They Run by Politicians or Bureaucrats?," Discussion Paper, University of Oxford 2024.

Freund, Caroline and Martha Denisse Pierola, "Export Superstars," *The Review of Economics and Statistics*, 12 2015, *97* (5), 1023–1032.

Gawande, Kishore, Pravin Krishna, and Marcelo Olarreaga, "A Political-Economic Account of Global Tariffs," *Economics & Politics*, 2015, *27* (2), 204–233.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text As Data," *Journal of Economic Literature*, 2019, *57* (3), 535–574.

Goldberg, Pinelopi K and Giovanni Maggi, "Protection for Sale: An Empirical Investigation," *The American Economic Review*, December 1999, *89* (5), 1135–1155.

__ and Nina Pavcnik, "The Effects of Trade Policy," in Kyle Bagwell and Robert Staiger, eds., *Handbook of Commercial Policy*, Vol. 1A, Amsterdam: Elsevier B.V. and North-Holland, 2016, chapter 3, pp. 161–206.

__ , Réka Juhász, Nathan J Lane, Giulia Lo Forte, and Jeff Thurk, "Industrial Policy in the Global Semiconductor Sector," Technical Report, National Bureau of Economic Research August 2024.

Goldstein, Harvey A, "The State and Local Industrial Policy Question: Introduction," *Journal of the American Planning Association*, 1986, *52* (3), 262–264.

Grimmer, Justin, Margaret E Roberts, and Brandon M Stewart, *Text As Data: A New Framework for Machine Learning and the Social Sciences*, Princeton, NJ: Princeton University Press, 2022.

Harrison, Ann and Andrés Rodríguez-Clare, "Trade, Foreign Investment, and Industrial Policy for Developing Countries," in Dani Rodrik and Mark Rosenzweig, eds., *Handbooks in Economics*, Vol. 5, Elsevier, 2010, chapter 63, pp. 4039–4214.

Hassan, Tarek A, Stephen Hollander, Laurence Van Lent, and Ahmed Tahoun, "Firm-Level Political Risk: Measurement and Effects," *Quarterly Journal of Economics*, 2019, *134* (4), 2135–2202.

Hausmann, Ricardo and Dani Rodrik, "Economic development as self-discovery," *Journal of development Economics*, 2003, *72* (2), 603–633.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd, "spaCy: Industrial-Strength Natural Language Processing in Python," 2020.

Hugging Face, "BERT Base Uncased Model," `https://huggingface.co/google-bert/bert-base-uncased` 2025. Accessed: 2025-01-10.

Johnson, Chalmers, *MITI and the Japanese Miracle: The Growth of Industrial Policy : 1925-1975*, Stanford, CA: Stanford University Press, 1982.

Johnson, Harry G, "Optimum Welfare and Maximum Revenue Tariffs," *The Review of Economic Studies*, August 1951, *19* (1), 28–35.

Juhász, Réka and Claudia Steinwender, "Industrial Policy and the Great Divergence," *Annual Review of Economics*, 2024, *16*.

_ and Nathan Lane, "The Political Economy of Industrial Policy," *Journal of Economic Perspectives*, 2024, *38* (4), 27–54.

_ , _ , and Dani Rodrik, "The New Economics of Industrial Policy," *Annual Review of Economics*, 2024, *16*.

Kalouptsidi, Myrto, "Detection and Impact of Industrial Subsidies: The Case of Chinese Shipbuilding," *The Review of Economic Studies*, April 2018, *85* (2), 1111–1158.

Kapczynski, Amy and Joel Michaels, "Administering a Democratic Industrial Policy," *Harvard Law and Policy Review*, 2023, *18* (2), 279–344.

Kee, Hiau Looi and Enze Xie, "Trade Policies Mix and Match: Theory, Evidence and the EU-Sino Electric Vehicle Disputes," Technical Report, World Bank Group, Washington, DC 2024.

Kim, Been, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, and Rory Sayres, "Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)," in Jennifer Dy and Andreas Krause, eds., *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80 of *Proceedings of Machine Learning Research* PMLR 10–15 Jul 2018, pp. 2668–2677.

Krippendorff, Klaus, *Content Analysis: An Introduction to Its Methdology*, 2nd ed., Thousand Oaks: Sage Publications Inc, 2004.

Krugman, Paul R. and Maurice Obstfeld, *International Economics: Theory and Policy*, 2nd ed., New York, NY: Harper Collins, 1991.

Lane, Nathan, "The New Empirics of Industrial Policy," *Journal of Industry, Competition and Trade*, 2020, *1* (2), 1–26.

_ , "Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea," Technical Report, Centre for the Study of African Economies 2021.

Lehmann, Sibylle H and Kevin H O'Rourke, "The Structure of Protection and Growth in the Late Nineteenth Century," *The Review of Economics and Statistics*, August 2011, *93* (2), 606–616.

Lin, Justin Yifu and Ha-Joon Chang, "Should Industrial Policy in Developing Countries Conform to Comparative Advantage or Defy it? A Debate Between Justin Lin and Ha-Joon Chang," *Development Policy Review*, 2009, *27* (5), 483–502.

Lindbeck, Assar, "Industrial Policy As an Issue in the Economic Environment," *The World Economy*, 1981, *4* (4), 391–406.

List, Frederick, *The National System of Political Economy*, Philadelphia, PA: J. B. Lippincott & Company, 1856.

Looi Kee, Hiau, Alessandro Nicita, and Marcelo Olarreaga, "Estimating Trade Restrictiveness Indices," *The Economic Journal*, jan 2009, *119* (534), 172–199.

Malouche, Mariem, José-Daniel Reyes, and Amir Fouad, "Making Trade Policy More Transparent: A New Database of Non-Tariff Measures," *World Bank Economic Premise*, November 2013, (128).

Nunn, Nathan and Daniel Trefler, "The Structure of Tariffs and Long-Term Growth," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 158–194.

Observatory of Economic Complexity, "Lithium oxide and hydroxide," `https://oec.world/en/profile/hs/lithium-oxide-and-hydroxide` 2025. Accessed: 2025-05-16.

OECD, "Measuring Distortions in International Markets: The Semiconductor Value Chain," Technical Report 234, OECD Publishing, Paris 2019.

_ , *OECD Inventory of Export Restrictions on Industrial Raw Materials 2024: Monitoring the use of export restrictions amid market and policy tensions*, Paris: OECD Publishing, 2024.

Pack, Howard and Kamal Saggi, "Is There a Case for Industrial Policy? A Critical Survey," *The World Bank Research Observer*, 2006, *21* (2), 267–297.

Passonneau, Rebecca J and Bob Carpenter, "The Benefits of a Model of Annotation," *Transactions of the Association for Computational Linguistics*, 2014, *2*, 311–326.

Pitelis, Christos N, "Industrial Policy: Perspectives, Experience, Issues," in Patrizio Bianchi and Sandrine Labory, eds., *International Handbook on Industrial Policy*, Cheltenham, UK: Edward Elgar Publishing, 2006, pp. 435–450.

Reidsma, Dennis and Jean Carletta, "Reliability Measurement without Limits," *Computational Linguistics*, September 2008, *34* (3), 319–326.

Rodrik, Dani, Princeton, NJ: Princeton University Press, 2008.

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky, "A Primer in BERTology: What we know about how BERT works," *Transactions of the Association for Computational Linguistics*, 2020, *8*, 842–866.

Romer, Christina D and David H Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–1084.

__ and __ , "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 2010, *100* (3), 763–801.

Stiglitz, Joseph E., Justin Yifu Lin, and Célestin Monga, "Introduction: The Rejuvenation of Industrial Policy," in Joseph E. Stiglitz, Justin Yifu Lin, and Célestin Monga, eds., *The Industrial Policy Revolution I: The Role of Government Beyond Ideology*, New York, NY: Palgrave Macmillan, 2013, pp. 1–15.

Taussig, Frank W, *The Tariff History of the United States*, New York, NY: GP Putnam's Sons, 1914.

Tenney, Ian, Dipanjan Das, and Ellie Pavlick, "BERT Rediscovers the Classical NLP Pipeline," in "Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics" Association for Computational Linguistics Florence, Italy July 2019, pp. 4593–4601.

Teti, Feodora A, "30 Years of Trade Policy: Evidence from 5.7 Billion Tariffs," Technical Report, ifo July 2020.

United States International Trade Commission, "Foreign Industrial Targeting and Its Effects on U.S. Industries Phase 1: Japan.," Technical Report, Washington, DC 1983.

Warwick, Ken, "Beyond Industrial Policy: Emerging Issues and New Trends," Technical Report, OECD, Paris 2013.

World Trade Organization, "World Trade Report 2006: Exploring the Links between the Subsidies, Trade and the WTO," 2006.

Online Appendix

Measuring Industrial Policy: A Text-Based Approach

Réka Juhász Nathan Lane Emily Oehlsen Verónica C. Pérez

A. Labeling: Annotator Agreement

Our classifier relies on labeled (annotated) data, which we use to train, validate, and test our classifiers. This appendix describes how we validate that human coders can consistently identify goals based on our definition and codebook.

A.1. Inter-coder reliability metrics

We find agreement between annotators in coding industrial policy classes (i.e., industrial policy, policies with other goals, and cases of not enough information). We assessed this consistency using two standard metrics, Krippendorff's alpha and Conger's kappa. Both metrics are suited for instances of more than two coders and take values between $[0 - 1]$, with 0 meaning perfect disagreement and 1 meaning perfect agreement.

Traditional content analysis considers Krippendorff's alpha values of 0.67-0.8 tolerable and above 0.8 high quality (Krippendorff, 2004). However, recent research questions these thresholds for machine learning applications, as intercoder reliability measures can be misleading (Reidsma and Carletta, 2008) or inapplicable. This is because if the source of disagreement is due to random noise, machine learning can tolerate data with lower agreement (Artstein, 2017; Passonneau and Carpenter, 2014).

However, if the disagreement is systematic, even reliability measures with values 0.80 and above will provide an unwanted pattern for the machine to detect (Reidsma and Carletta, 2008). Therefore, we treat these metrics as general guidance, considering our measures roughly reliable, particularly in annotation rounds 2-4, where Krippendorff's alpha approaches 0.8.

B. Model: Core LLM (BERT) and Benchmark (Logistic)

This technical appendix describes the models used in our analysis and describes their parameters, tuning procedures, and training workflows. Our primary model is BERT (Devlin et al., 2019) (Bidirectional Encoder Representations from Transformers), a deep neural network pre-trained on large-scale natural language corpora.

In our application, as in many others, we fine-tune the pre-trained BERT model on a task-specific dataset to perform our custom classification task. To benchmark performance in a more transparent and interpretable manner, we compare BERT to a regularized logistic regression classifier. For this logistic classifier, documents are represented using unigrams and bigrams, which are vectorized via term frequency–inverse document frequency (TF–IDF). We use the best-performing variant of logistic regression, which uses L_1 regularization.

B.1. Fine-Tuned BERT Model

1. BERT: Model Overview. For our BERT-based classifier, we use the `bert-base-uncased` model (Hugging Face, 2025), which we then fine-tuned for our specific three-class classification task. This pre-trained model consists of 12 transformer encoder layers, 768 hidden units, and 12 attention heads. To adapt it to our task, we append a randomly initialized linear classification layer atop the pooled output of the final hidden layer, with an output dimension corresponding to the three target classes: industrial policy, not industrial policy, and not enough information.

Because BERT is context-aware, text inputs require only minimal pre-processing. Unlike traditional NLP pipelines (e.g., stop-word removal, lemmatization), BERT relies on its built-in tokenizer, WordPiece, to convert inputs into subword tokens. This preserves much of the original text structure and avoids additional filtering steps typically used in logistic classifiers.

We fine-tune and evaluate our BERT model using the annotated data splits: $\mathcal{D}_{\text{train}}$, $\mathcal{D}_{\text{val}}$, and $\mathcal{D}_{\text{test}}$ (training, validation, and test sets, respectively). These splits consist of 3,384 samples in $\mathcal{D}_{\text{train}}$, 587 in $\mathcal{D}_{\text{val}}$, and 440 in $\mathcal{D}_{\text{test}}$, with labels assigned based on our predefined annotation procedure.

The full implementation—including fine-tuning and hyperparameter optimization—was conducted using the Hugging Face Transformers library within a PyTorch environment. This setup enabled efficient loading of the pre-trained `bert-base-uncased` model and tokenizer, definition of the sequence classification head, and execution of training and evaluation loops during the hyperparameter search.

Note that the training process for a complex language model like BERT is not deterministic. The process involved in creating a fine-tuned BERT model involves randomness. Even with global seeds, randomness comes from differences in Python libraries, GPU processing, data shuffling, and differences in initialization weights. Thus, while the outputs of the classifier may have deterministic behavior, some aspects of training introduce randomness.

2. BERT: Hyperparameter Tuning and Training. Before training our BERT model, we first performed hyperparameter tuning using the training and validation sets

($\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{val}}$). We then followed standard practice by training the final model on the combined $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$ split using the optimal hyperparameters. To address class imbalance, we use standard oversampling throughout the hypertuning experiment and final training.

All experiments—hyperparameter tuning and final training—were conducted on an NVIDIA GH200 480GB GPU using PyTorch version 2.6.0 and Hugging Face Transformers version 4.50.3. Hyperparameter tuning was carried out with Optuna version 4.2.1. To ensure reproducibility, we set a global random seed across Python's random module, NumPy, and PyTorch (including CUDA); individual Optuna trials also used derived seeds. Mixed-precision training with bfloat16 was employed for computational efficiency. Data loading was optimized using multiple worker threads (up to 63 on our GPU setup) and pinned memory. We used the `bert-base-uncased` tokenizer throughout.

The hyperparameters varied during optimization included learning rate, batch size, number of training epochs, and weight decay; see Appendix Table B.1. We used the standard `AdamW` optimizer with a fixed warm-up phase (warm-up ratio of 0.06) and the default hidden dropout probability optimized for BERT (0.1).

Table B.1: Hyperparameter Search Space for BERT Model Tuning

Hyperparameter	**Range / Values**	**Sampling Strategy / Type**	**Best Parameter**
Learning Rate	[1e-5, 8e-5]	Log-uniform	6.0593×10^{-5}
Batch Size	{4, 8, 16, 32}	Categorical	32
Number of Train Epochs	[2, 5]	Integer (Uniform)	4
Weight Decay	[1e-5, 1e-3]	Log-uniform	3.0229×10^{-6}

Hyperparameter optimization was conducted using an implementation of the Bayesian Tree-structured Parzen Estimator (TPE) algorithm (Bergstra et al., 2011; Akiba et al., 2019).[1] The TPE sampler was configured to maximize the macro F1-score on the validation set $\mathcal{D}_{\text{val}}$, with early stopping implemented via a median pruner to reduce runtime.

We executed 150 trials over the hyperparameter space defined in Appendix Table B.1. Large neural networks like BERT can exhibit variance in performance across runs, even with the same hyperparameter configuration. For this reason, each trial was replicated three times ($N = 3$), and the mean F1-score was used to evaluate each configuration.

1. We use Optuna's implementation of TPE, a Bayesian optimization method that efficiently explores the hyperparameter space by learning from prior evaluations. This makes it especially well-suited for tuning computationally intensive models like BERT.

The final model was trained using the best-performing hyperparameter set, applied to the combined $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$ dataset. As of our current model `v1`,[2] the selected hyperparameter values are presented in Appendix Table B.1.

We assess the performance of the final model on the held-out test set $\mathcal{D}_{\text{test}}$. For completeness, Appendix Table H.8 reports predictive performance for the three-class BERT model on the test set ($\mathcal{D}_{\text{test}}$), as well as on the validation ($\mathcal{D}_{\text{val}}$) and training ($\mathcal{D}_{\text{train}}$) splits. This table is an expanded version of the main table shown in the text.

B.2. Logistic Classifier Benchmark

We use a three-class logistic regression classifier, based on TF–IDF vectorization, as a transparent benchmark throughout the paper. Specifically, we report results using the logistic model with L_1 regularization. This benchmark model uses unigrams and bigrams as input features and applies Lasso (L_1) regularization with a strength of $C = 0.09$ (equivalently, a regularization strength of $\lambda = 11.1$).

Given our setting and large vocabulary, the "more sparse" L_1-penalized model outperforms logistic regressions using Ridge (L_2), and performs slightly better than Elastic Net. We describe the full pipeline, model selection process, and resulting performance below.

1. Logistic: Text Processing, Tokenization, and Vectorization. Text is processed and represented differently in our baseline logistic classifier compared to the BERT-based model. In the logistic pipeline, each policy description is converted into a numerical array using standard Term Frequency–Inverse Document Frequency (TF–IDF) representations, incorporating both unigrams and bigrams.

Text is first tokenized and cleaned using the Python-based NLP library spaCy (version 3.8.3).[3] The preprocessing pipeline follows standard conventions to reduce noise, normalize linguistic structure, and enhance classifier performance. This involves several steps: tokens identified by spaCy as punctuation or number-like are removed; the remaining tokens are lemmatized (i.e., converted to their base morphological form), transformed to lowercase, and filtered through a stopword list. This stop list combines spaCy's default English stopwords with domain-specific terms, including units, to remove non-informative content. Where possible, we use spaCy's ready-made libraries for our cleaning and filtering (e.g., stop word lists and lemmatizers). Tokens that become empty during preprocessing are discarded, and the pipeline delivers a cleaned and lemmatized sequence for vectorization.

Processed text is then vectorized using a TF–IDF vectorizer from the `Scikit-learn` library. The vectorizer is configured with max = 95% to exclude overly common

2. `https://huggingface.co/industrialpolicygroup/industrialpolicy-classifier`
3. See Honnibal, Montani, Van Landeghem and Boyd (2020).

terms and min = 2 (minimum 2 observations) to remove rare ones. Sublinear term frequency scaling is applied, along with the default L_2 normalization.

Following vectorization, a standard scaler is applied to further normalize the feature representation for logistic classification.[4] This step is particularly important for regularized models such as Lasso and Elastic Net (with Saga solvers), which are sensitive to the magnitude of input features. Scaling ensures that all features contribute proportionately to the regularization penalty, preventing features with larger inherent values from disproportionately influencing model coefficients. This results in more stable performance and more interpretable model outputs.

2. Logistic: Hyperparameter Tuning, Training, and Model Selection. We select the tuned logistic model—Ridge (L_2), Lasso (L_1), and Elastic Net—which performs the best out of sample. Below, we describe the hyperparameter tuning process, the training, and the out-of-sample performance of each.

Table B.2: Logistic Variants and Hyperparameter Tuning

Logistic Model	Hyperparameter Space	Solver	Scaled Text	Best F1	Best Hyperparameter
Ridge (L_2)	$C \in [.25, 2]$	lbfgs	No	0.8403	$C = 0.5$
Ridge (L_2) Scaled	$C \in [.05, 1]$	saga	Yes	0.8231	$C = 0.25$
Lasso (L_1)	$C \in [.075, 2]$	saga	Yes	0.8513	$C = 0.09$
Elastic Net	$C \in [.05, 1]$	saga	Yes	0.8536	$C = 0.075$
	L_1 ratio $\in [.5, .99]$				$\frac{L_1}{L_2} = 0.95$

Notes: All F1-scores are macro-averaged and computed as the mean across all k-folds of the validation set. Scaling refers to the application of a standard scaler to text features after TF–IDF vectorization; means are not subtracted to preserve sparsity in the TF–IDF representation. The solver indicates the optimization algorithm used to estimate the logistic regression parameters: `lbfgs` (Limited-memory Broyden–Fletcher–Goldfarb–Shanno), a quasi-Newton method effective for L_2 regularization; and `saga`, a stochastic optimization algorithm suitable for L_1 and Elastic Net penalties.

We use a standard grid search algorithm to optimize the hyperparameters of our logistic regression models implemented using Scikit-learn's `GridSearchCV`. For each variant, we select the logistic classifier that yields the highest macro-averaged F1-score—average F1 for validation sets across folds—during the tuning process.

We began by identifying key hyperparameters for each model, such as the regularization strength C and the L1 ratio (for Elastic Net), and defined a discrete set of candidate values Λ_i for each. The resulting grid $\mathcal{G}$ comprised all possible combinations of these values.[5]

4. The `Scikit-learn` standard scaler scales features to unit variance by dividing by their standard deviation. To preserve sparsity in the TF–IDF representation, the mean is not subtracted.
5. The Cartesian product $\Lambda_1 \times \cdots \times \Lambda_k$.

For every hyperparameter combination $h \in \mathcal{G}$, we trained a complete logistic model pipeline. This pipeline included text preprocessing, TF–IDF vectorization to produce features X, optional feature scaling, and logistic regression using the classifier $g_h(X'; w)$ parameterized by h.

Each candidate model M_h was evaluated using a separate, fixed validation set $\mathcal{D}_{\text{val}}$, while the training set $\mathcal{D}_{\text{train}}$ was used solely to estimate model parameters w. We built on standard k-fold cross-validation by applying a predefined split: each M_h was trained on the full $\mathcal{D}_{\text{train}}$ to estimate $\hat{w}_h$, and its performance was evaluated using the macro-averaged F1-score, denoted $\text{F1}_{\text{macro}}(M_h(\mathcal{D}_{\text{val}}))$. Internally, `GridSearchCV` applied a three-fold split of $\mathcal{D}_{\text{val}}$ during this evaluation.

We then selected the hyperparameter configuration h^* that maximized validation performance.[6] This grid search and model selection procedure was repeated independently for each regularized logistic regression variant. The resulting best model M_{h^*}, trained using the optimal configuration h^*, was retained for final evaluation.

Finally, we selected the logistic variant with the highest out-of-sample performance on the test set $\mathcal{D}_{\text{test}}$, reported in Appendix Table B.3. Specifically, each optimized variant from the grid search (Appendix Table B.2) was retrained on the combined $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$ split and evaluated on the held-out test set $\mathcal{D}_{\text{test}}$. Appendix Table B.3 reports the out-of-sample performance and shows that the Lasso model performs best. The tuned Lasso model with $C = 0.09$ outperforms both Ridge and Elastic Net. For completeness, Appendix Table B.3 also reports results from a non-optimized logistic regression model using default L_2 strength parameters for the logistic regression library, which compares our tuned estimates to a standard, out-of-the-box baseline.[7]

6. Formally, the optimization problem:

$$h^* = \arg\max_{h \in \mathcal{G}} \ \text{F1}_{\text{macro}}\left(M_h(\mathcal{D}_{\text{val}})\right).$$

7. Default $C = 1$ in Scikit-learn's Ridge logistic classifier.

Table B.3: Predictive Performance of Different Logistic Models

Logistic Model	Class	Precision	Recall	F1-score
No Optimization	Industrial Policy	0.809	0.839	0.824
	No Industrial Policy	0.848	0.787	0.816
	Not Enough Information	0.897	0.903	0.900
	Accuracy			0.868
	Macro Avg	0.851	0.843	0.847
	Model Average	0.868	0.868	0.867
Ridge	Industrial Policy	0.820	0.875	0.847
	No Industrial Policy	0.889	0.842	0.865
	Not Enough Information	0.922	0.912	0.917
	Accuracy			0.891
	Macro Avg	0.877	0.876	0.876
	Model Average	0.892	0.891	0.891
Lasso	Industrial Policy	0.867	0.875	0.871
	No Industrial Policy	0.971	0.882	0.924
	Not Enough Information	0.921	0.942	0.932
	Accuracy			0.916
	Macro Avg	0.920	0.900	0.909
	Model Average	0.917	0.916	0.916
Elastic Net	Industrial Policy	0.858	0.875	0.867
	No Industrial Policy	0.971	0.882	0.924
	Not Enough Information	0.921	0.938	0.930
	Accuracy			0.914
	Macro Avg	0.917	0.898	0.907
	Model Average	0.915	0.914	0.914

Notes: This table presents the predictive performance of the logistic models on the test data split. Each model is trained using optimal hyperparameters on the train/validation splits. For each model, we detail standard classification metrics—Precision, Recall, and F1-Score—for each of the three classes. Accuracy is given for the model's predictions across all classes. The classes are 'IP Goal', 'No IP Goal', and 'Not Enough Information'. Additionally, the table shows the Macro Average and Weighted Average for these metrics across classes.

C. BERT Validation and Concept Activation Vectors

This section outlines our workflow for Testing with Concept Activation Vectors (TCAV), including key implementation details. We use TCAV to interpret the internal reasoning of our fine-tuned BERT model by quantitatively assessing its sensitivity to a high-level, human-understandable concept—in this case, the concept of an "industrial policy goal," which we denote as C.

The following implementation was carried out in Python 3.9 using several core libraries, including `PyTorch (v2.5.1)`, `Hugging Face Transformers (v4.35)`, `Captum (v0.8.0)`, `Scikit-learn (v1.2.0)`, and `TensorFlow (v2.19.0)`. This analysis was performed on a high-performance computing node equipped with an NVIDIA GH200 GPU (480 GB).

C.1. Concept and Example Preparation

This section describes the construction of sets used for TCAV analysis. The concept C is defined through a curated collection of 300 representative example texts that reflect the notion of an industrial policy goal. To establish a rigorous baseline, we also compiled a diverse set of 300 negative examples, drawn from 16 distinct conceptual categories unrelated to industrial policy. All texts were tokenized using BERT's WordPiece tokenizer, with sequences truncated or padded to a maximum length of 128 tokens.

1. Concept Generation. Testing with Concept Activation Vectors (TCAV; Kim *et al.*, 2018) assesses the influence of human-interpretable concepts on neural network predictions. Rather than probing individual input features, TCAV learns linear classifiers in a model's activation space to separate concept examples from random counterexamples. When a concept consistently aligns along a particular direction in this high-dimensional space, it signals that the model relies on that concept.

The list C of industrial policy goals consists of 300 examples spanning six thematic categories designed to capture variation across domains. These include manufacturing and industrial goals—such as enhancing domestic production and developing advanced manufacturing—as well as objectives related to the service economy, the digital economy, and the agricultural economy. For example, agricultural goals involve developing horticultural exports and boosting fruit production. Additional targets focus on produce and horticulture, including efforts to drive innovation, diversify crops, and optimize post-harvest value addition. Finally, international trade promotion objectives aim to expand export markets, improve market access, and strengthen trade facilitation infrastructure.

We generated our concept list using our industrial policy definition, seed examples, and generative AI (OpenAI's GPT-4o and Anthropic's Claude 3.7 Sonnet). We began by curating seed examples that provided the semantic backbone for each thematic category and ensured that our concept definitions were anchored in real-world language. We then populated the list, using generative LLMs with prompts that preserved the institutional tone of official documents while introducing lexical and syntactic variety. We used generative LLMs to enrich the set with alternative verb choices, varied terminology, and new phrasing, all while maintaining the policy-oriented framing of each goal.

We deliberately diversified sentence form: imperatives ("Strengthen national production capacity..."), infinitive clauses ("To enhance supply-chain resilience..."), participial phrases ("Aimed at fostering advanced manufacturing..."), and complex purpose statements ("With the objective of promoting digital infrastructure..."). This variation ensures that TCAV interprets semantic content rather than picking up on repetitive syntactic patterns.

Once examples were generated, each candidate was reviewed for semantic relevance, linguistic authenticity, and distinctiveness from other concepts, eliminating any ambiguous or overlapping entries. Finally, a rigorous quality-assurance pass removed duplicates, confirmed balanced coverage across all six thematic areas, and verified that every example retained the technical characteristics of declarative, formal policy writing.

2. Negative (Random) Example Generation.

Main Negative Sample Set (In-Distribution). To estimate the concept activation vector, we create a diverse pool of negative (or random) text examples clearly distinguishable from the target concept. These negative examples exclude economic semantics while maintaining the surface form (formal, declarative English) of concept C.

We follow the literature and use random, neutral text that does not contain concept semantics. Specifically, we create sixteen sets, each containing 300 random strings, $\mathcal{N}_1, \ldots, \mathcal{N}_{16}$, drawn in-distribution from the GTA dataset. We sample distinct, random sentences (and fragments) that do not contain concept vocabulary and semantics (industrial policy goals).

Robustness: Out-of-Distribution Negative Sets. For robustness, we also draw sixteen sets of random sentences $\mathcal{N}'_1, \ldots, \mathcal{N}'_{16}$ from out-of-distribution sources. Specifically, we use LLMs to sample from public corpora (Wikipedia articles, arXiv abstracts, PubMed papers, Stack Overflow, Common Crawl news). We constructed sixteen thematically distinct negative sets, choosing topics unrelated to industrial or economic policy: legislative history, personal finance, clinical psychology, natural sciences, physics, earth science, sleep, diplomacy, climate change, justice, demography, cul-

tural preservation, electoral studies, health events, geopolitical news, and research objectives.[8]

Our generation process employed explicit constraints on semantic content related to industrial policy goals, prohibiting policy-related vocabulary (e.g., "boost," "enhance," "strategic," "development") and goal-oriented syntactic patterns (e.g., infinitive constructions beginning with "To boost"). Each LLM call returned approximately 300 sentences, formatted as comma-separated tables with columns: Topic, ID, Category, and Sentence.

Negative sentences were systematically generated to match linguistic parameters of concept set C.[9] We targeted sentence lengths between 6 and 16 words ($\mu \approx 11$), with each LLM verifying this distribution post hoc. Our generation script enforced syntactic variety by requiring at least 15% fragments, 20% subordinate-clause sentences, and Shannon lexical entropy within ±0.05 of the positive corpus average.

Finally, each string underwent validation to ensure no conceptual overlap with industrial policy themes while preserving the linguistic structure necessary for fair TCAV analysis. During processing, we employed a custom script to confirm that each topic set excluded predefined economic keywords.

C.2. Activation Extraction and Pooling

Text inputs are tokenized and fed through our frozen fine-tuned BERT-base classifier (weights are no longer updated). We extract hidden-state activations from the later transformer layers. These produce activation tensors with dimensions `(128, 768)` for each layer.

We then pool these high-dimensional activations, condensing each activation tensor into a single vector that summarizes information for the entire input text. While each input generates an activation vector for every token (a tensor of shape `(128, 768)`), TCAV requires one fixed-size vector per input text. Pooling aggregates the tensor to a single `768`-dimensional vector for each input, directly connecting conceptual information and model predictions at the sequence level.

We consider two pooling strategies:

i. [CLS] Pooling. Our primary strategy uses `[CLS]` token pooling, naturally, given that the `[CLS]` token plays a central role in BERT predictions. The `[CLS]` token is a special symbol that BERT automatically adds to the start of every input text.

8. Topics are purposefully diverse. Themes like personal finance may be proximate to economics—entailing personal economic activity such as purchases and savings—but remain distinct from industrial policy. Others, like natural sciences, are unrelated to social science or human activity. For each negative topic, we sample across subtopics.

9. The target concept set has the following characteristics: average string length ($\mu \approx 11$ words), syntactic patterns (30% infinitive constructions, 20% prepositional phrases), and vocabulary register (formal policy language with goal-oriented terminology).

During training, BERT learns to use this token's final hidden state to capture the overall meaning or "summary" of the input. Instead of aggregating all tokens, we take the activation vector of just the `[CLS]` token. This single vector serves as a compact, sequence-level representation commonly used for classification tasks and interpretability analyses.

ii. Mean-Pooling Robustness. For robustness, we perform a separate analysis with mean pooling. This method averages the activation vectors of all tokens in the input sequence (including special tokens like `[CLS]` and `[SEP]`). By taking the mean across all token activations, we obtain a single vector that captures information distributed throughout the sequence. This provides an alternative, more holistic sequence-level representation, helping ensure our findings are not overly dependent on `[CLS]` token properties alone.

C.3. Extracting Concept Activation Vector (CAV) Using Logistic Regression

A Concept Activation Vector (CAV) is a unit vector $\mathbf{v}_C$ in the model's activation space that points in the direction most reliably encoding concept C. Formally, let $A_C = \{a_1, a_2, \ldots, a_m\}$ be the set of pooled activation vectors corresponding to positive concept examples, and let $A_N = \{a'_1, a'_2, \ldots, a'_n\}$ be the set corresponding to negative examples. We use a linear binary classifier to distinguish between these sets and extract $\mathbf{v}_C$.

To derive a CAV, $\mathbf{v}_C$, we train an L_2-regularized logistic regression classifier on A_C and A_N. The pipeline includes normalization using `StandardScaler`. Due to the high dimensionality of the activation space (768 features), we incorporate PCA to reduce dimensionality before fitting the classifier. The classifier is trained and evaluated using K-fold cross-validation ($K = 6$ folds). The logistic regression classifier robustly distinguishes the activations (mean accuracy 98%).[10]

The CAV is simply the normalized coefficient vector from logistic regression. Formally, the learned hyperplane separating A_C and A_N is defined by $\mathbf{w}^\top \mathbf{a} + b = 0$, where $\mathbf{w}$ is the coefficient vector, $\mathbf{a}$ is the activation vector, and b is the bias term. The CAV, $\mathbf{v}_C$, is the unit vector pointing in the direction of coefficients $\mathbf{w}$: $\mathbf{v}_C = \frac{\mathbf{w}}{\|\mathbf{w}\|}$, representing the direction orthogonal to the decision boundary that maximally increases the model's confidence in identifying activations corresponding to concept C.

10. We use the PCA components to back-project (re-inflate) the classifier coefficients to the original activation space and correct for scaling.

C.4. Calculating TCAV

The final stage quantifies concept C's influence on the model's prediction for each class k. We use the Concept Activation Vector $\mathbf{v}_C$ to calculate the TCAV score, which measures how important the concept is for predicting that class.

1. TCAV Score and Calculating Sensitivity. This score provides a quantitative measure of the concept's global importance for classifying inputs as class k. The TCAV score for concept C and class k is the fraction of examples in the evaluation set $\mathcal{X}_k$ with positive conceptual sensitivity:

$$\text{TCAV}_{C,k} = \frac{1}{|X_k|} \sum_{x \in X_k} \mathbb{I}\big[S_{C,k}(x) > 0\big], \tag{C.1}$$

where $\mathcal{X}_k$ contains examples from class k in our evaluation dataset. The key term, $S_{C,k}(x)$, is the sensitivity of the model's prediction for class k with respect to activations generated by input x. When $S_{C,k}(x) > 0$, moving input x's activation vector toward the concept increases the likelihood of classification as class k.

Computing equation C.1 requires calculating the "conceptual sensitivity," $S_{C,k}(x)$:

$$S_{C,k}(x) = \nabla h_k(f(\mathbf{x})) \cdot \mathbf{v}_C, \tag{C.2}$$

where h_k maps activations to the prediction logit for class k. The gradient $\nabla h_k(f(\mathbf{x}))$ shows how the logit changes with respect to the activation vector. However, calculating gradients in deep networks like BERT is problematic due to nonlinearities from activation functions. While these nonlinearities enable learning complex patterns, they make gradients unreliable (noisy, unstable, and variable).

To overcome this limitation, we employ Integrated Gradients (IG), a more robust method for attribution. Sensitivity is calculated as follows,

$$S_{C,k}(x) = \text{IG}_k(\mathbf{x}) \cdot \mathbf{v}_C, \tag{C.3}$$

where $\text{IG}_k(\mathbf{x})$ is a vector of Integrated Gradients: the path integral of gradients for the model's class k prediction with respect to layer activations for input $\mathbf{x}$.

2. Integrated Gradients and Baseline Choice. Integrated Gradients provides a robust attribution score, computed along a straight-line path from baseline activation $f(\mathbf{x}')$ to input activation $f(\mathbf{x})$.[11] Here, $\mathbf{x}'$ represents a neutral baseline input, such

11. Formally, the continuous integral:

$$\text{IG}_k(\mathbf{x}) = (f(\mathbf{x}) - f(\mathbf{x}')) \times \int_{\alpha=0}^{1} \nabla h_k(f(\mathbf{x}') + \alpha(f(\mathbf{x}) - f(\mathbf{x}')))d\alpha. \tag{C.4}$$

as [MASK] tokens, and α is an interpolation constant for the path. By integrating gradients along this path, IG ensures attribution is fully distributed among features, providing a faithful sensitivity measure not susceptible to gradient saturation.

The continuous integral must be approximated numerically. Our workflow uses a Riemann sum with `n_steps` set to 100. For each input, the path from baseline to input activations is divided into 100 discrete intervals. The gradient is computed at each point and averaged to approximate the integral. This provides an efficient approximation of the integral.

A critical component of Integrated Gradients is choosing the baseline a', which should represent an uninformative input. In our workflow, this baseline is derived from an input sequence where every token has been replaced by the special [MASK] token.[12]

This [MASK] baseline is well-suited for BERT because the model was pre-trained to understand this token as a placeholder for unknown information. Unlike a simple zero vector, the [MASK] baseline is an "in-distribution" neutral input that preserves positional embedding information. This ensures that IG attribution measures the influence of text content (token identities) without confounding from absent positional cues or baselines far from the model's learned distribution.

3. Sampling Example Texts from Final Prediction Dataset X_k. TCAV requires us to use samples for each category from our "target dataset"—our final prediction dataset. Before calculating gradients, we select a representative sample of texts, X_k, for each output class k.

We use a confidence-based approach to identify high-quality text examples for concept analysis, sampling texts that unambiguously belong to each class k. We focus on predictions above the 90th percentile of logits, creating a pool of high-confidence candidates from which we randomly sample 350 examples. This approach is valuable for downstream TCAV tasks, where training example quality may impact concept detection and interpretation reliability.

12. We construct it by creating an input sequence of the same length as the original input (e.g., 128 tokens) with every token replaced by [MASK] from the model's vocabulary.

D. Validating the GTA using OECD data on Export Restrictions on Industrial Raw Materials

This section contains a detailed description of the GTA data validation exercise referenced in Sections 3 and 6.C.[13] We first describe the OECD dataset. Next, we explain our hand-matching protocol. Then, we discuss the findings.

The OECD's inventory lists export controls from 2009-2021 on 65 industrial primary commodities across metals, minerals, and wood. Policies are verified using official government sources (OECD, 2024). The OECD covers the 80 countries that are significant producers of any of these products. For each country, coverage is limited to the subset of products for which that country is a significant producer.

To understand the quality of the Global Trade Alert in terms of its ability to enumerate relevant policies, we hand-match policies across the two data sources. To do so, we need to transform and filter the data to render them comparable. This process involves a few steps. First, the OECD reports the stock of policies annually, whereas the GTA enumerates only new policies–i.e., it reports the flow. We thus transform the OECD data to also be defined in terms of annual flows.

Second, we need to filter both data sources for only those that fall under both organizations' remit. We count GTA policies as in the OECD domain if they 1) are in place at any point between 2009 and 2021, 2) have a listed "Measure Type" of "Export ban," "Export licensing requirement," "Export quota," or "Export tax," 3) affect products that the OECD covers for that country, and 4) are backed up by an official government source provided by the GTA. This provides a lower bound of GTA policies in the OECD domain as we do not include GTA measure types that may span both export controls (within the OECD remit) and other export policies (not in the OECD remit), such as "Export-related non-tariff measure."

We mark an OECD policy as in the GTA domain if it is 1) introduced after November 2008, 2) implemented unilaterally, and 3) there is some change, no matter how small, from the previous policy.[14] Our estimate of OECD policies in the GTA domain is an upper bound. The key reason is that we include policies with only minimal changes from previous policies. These policies are unlikely to meaningfully

13. We are grateful to our research assistant, Lottie Field, for her extensive work in constructing a comparable version of the OECD and GTA datasets and meticulously hand-matching them over many months.

14. We also exclude OECD policies which we cannot identify using the information provided by the OECD. For example, for Mexico the only details on individual policies that the OECD provides are the date of introduction, type of policy instrument and affected sectors. This information was not sufficient to identify specific policies.

affect global trade flows, so we likely include many OECD policies that do not meet the GTA's reporting thresholds.[15]

We define two types of matched policies. An OECD policy has a full GTA match if we can pinpoint the same policy document for both entries (e.g., for Zambia, both the OECD and GTA list Statutory Instrument No. 40 of 2020 suspending the export duty on precious metals). A GTA policy partially matches an OECD policy if it 1) uses the same policy instrument, 2) affects the same industrial primary commodities, and 3) is announced within one year of the OECD policy being introduced. We think of GTA policies that partially match an OECD policy as being in the same "policy series." Generally, these are policies that precede, replace or amend the OECD policy. A partial match indicates that the GTA covers policies in the same area, even if it does not capture the exact policy.[16]

Given the large size of the data, we hand-matched the policies to a random subsample of countries stratified by income. The countries are as follows: Ethiopia, Rwanda, Guinea, Zimbabwe, Zambia, India, Kenya, Laos, Vietnam, Ukraine, Philippines, Indonesia, Angola, Egypt, Guatemala, Tunisia, Thailand, Colombia, Botswana, Turkey, Brazil, Oman, UAE, Canada.[17]

Appendix Figure G.15 shows that, in general, there is a fair amount of overlap across the two datasets. Of the policies identified in the OECD dataset, 36% have an exact match in the GTA, and 65% have a partial match. This is a conservative lower bound mainly for the reason (noted above) that many policies enumerated by the OECD are minor policy changes that do not satisfy the GTA's criteria of affecting global trade flows in a meaningful way. Similarly, 63% of the relevant policies in the GTA can be exactly matched to an OECD policy. That is, despite the much narrower focus of the OECD data collection effort, a meaningful share of relevant policies are not identified by the OECD, but *are* identified by the GTA.

15. For example, the OECD records when Brazil reduced its export quota on Lithium oxide from 50 to 10 metric tons in 2016. According to the Observatory of Economic Complexity (2025), Brazil is a negligible exporter in this category, so this small change in the export quota is unlikely to affect trade flows, and as such, will not be enumerated by the GTA.
16. For example, for Indonesia, the OECD lists a January 2009 Ministry of Trade regulation mandating domestic letters of credit to export certain goods. This partially matches with the GTA listing of the Ministry of Trade's March 2009 policy, which updates the earlier policy by, *inter alia*, specifying that the requirement only applies to exports worth over 1 million US dollars.
17. Our initial sample also included Mongolia, Peru, Mexico, Australia, the U.S. and Qatar. We were unable to perform hand-matching for these countries because none of the policies listed by the OECD were in the domain of the GTA.

E. Robustness of Measures

Our descriptive analysis presents results in three distinct ways. We do so to address variations in reporting granularity and aggregation across countries, policies, and other factors in the GTA source data. Here, we provide a concrete example that illustrates how different reporting standards across countries may bias measurement based on raw policy counts (the baseline measure).

Consider how GTA enumerates the US EXIM Bank's disbursements of support to firms. A typical policy enacted by the US EXIM Bank reads as follows in the GTA: "In March 2013, the Export-Import Bank of the United States (EXIM) provided a guarantee for a working capital loan given to Gaffney-Kroese Electrical Supply Corp." Source: GTA. Now consider how GTA enumerates the Indonesian Export-Financing Agency's disbursements to firms: "On 14 July 2015, the Indonesian Finance Ministry announced regulation 134/PMK.08/2015 allowing the Indonesian Export-Financing Agency LPEI to support export-oriented Indonesian companies through Special Assignments from the Finance Ministry." Source: GTA.

In this example, a comparison of policy counts between the US and Indonesia will overestimate industrial policy activity in the US, as individual disbursements for firms are enumerated as separate policies, whereas in Indonesia, the new support package is counted as one policy. This is a well-known issue with the enumeration of policies in the GTA (Evenett, 2019). It leads to a concern that our baseline measure of industrial policy may overestimate activity in countries with higher administrative capacity or greater government transparency.

In the main text, we deal with this challenge by conducting the analysis using three different measures of industrial policy activity. The second measure, which uses only national policies, drops all policies implemented at the firm level. In our example above, this would imply dropping the US policy, but keeping the Indonesian one. The third measure, which enumerates all policies at the implementing agency level, would retain information from both the examples listed above, but the US EXIM Bank and the Indonesian Export-Financing Agency are enumerated only once (or once per year, or once per policy instrument, depending on the analysis), no matter how many distinct GTA policies they appear in.

F. Agencies Implementing Industrial Policy

Below, we describe how we extract data on "implementing agencies" from Global Trade Alert policy descriptions. The following account closely follows the appendix of Juhász and Lane (2024). Field (2024) developed this method by extracting

implementing agency-level data from the GTA. We apply this process to the entire GTA dataset.

We use OpenAI's ChatGPT to extract agencies administering and deploying industrial policy from textual policy descriptions. Our workflow is algorithmic and leverages ChatGPT's API (Application Programming Interface) for the specialized task of extracting industrial policy institutions and country references from unstructured text. This workflow, implemented in Python, was executed in November 2023.

Specifically, we extract agencies implementing industrial policy using ChatGPT 3.5 (GPT-3.5-turbo-0613), which we fine-tuned to identify implementing agencies and country names from policy summaries. First, we select test and training samples from the source dataset. Next, we develop a custom prompt instructing ChatGPT on mining implementing agencies from policy text, integrating expected replies into the training sample. This labeled data specifies how ChatGPT should extract the implementing agencies and the expected output.

Third, we use the processed training data to fine-tune the baseline GPT-3.5-turbo-0613 model through the OpenAI API. The labeled data is fed into a fine-tuning pipeline that updates the model's weights to handle the custom implementing agency extraction task. Fourth, we evaluate and validate the fine-tuned model using the test sample.

Finally, we deploy the fine-tuned model to extract agencies implementing industrial policy from the entire source dataset. The extracted implementing agencies are processed, cleaned, and validated manually, generating a comprehensive dataset of agencies implementing industrial policy from the source data.

We manually clean the extracted implementing agency names with two key steps. First, we standardize the implementing agency names to ensure consistency across different spellings and forms. For example, the Italian development bank "Cassa Depositi e Prestiti" may appear as "Cassa Depositi e Prestiti," its abbreviation "CDP," or as "Italian National Development Bank." Second, we remove non-public implementing agencies, such as private companies, based on shareholder composition or company history, using the method developed by Field (2024). We also remove regional and supranational organizations.

G. Appendix Figures

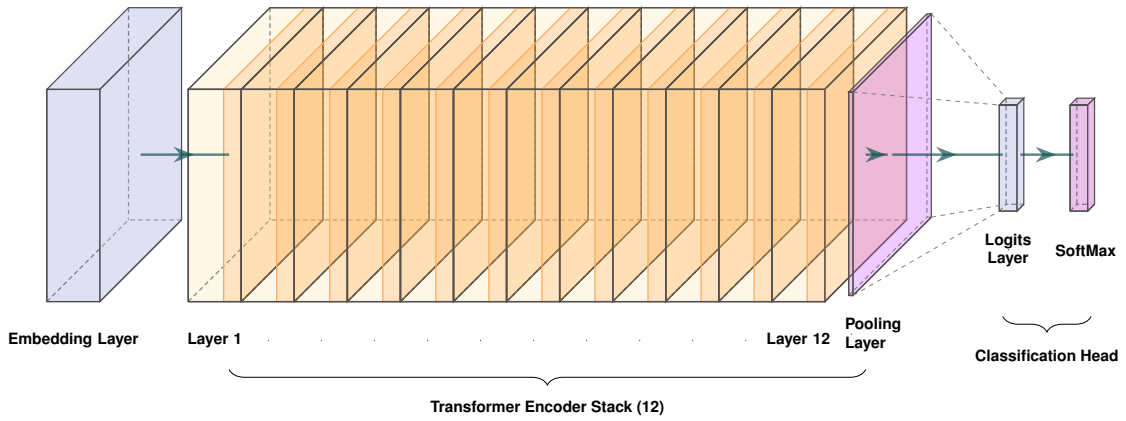


Figure G.1: Fine-Tuned BERT-based Classifier Architecture

Notes: This figure illustrates the architecture of a fine-tuned BERT classifier (three-class). Input text is first converted into high-dimensional numerical representations by the embedding layer. These embeddings are then processed by BERT's transformer encoder stack, which consists of 12 layers that sequentially build a deep, contextualized understanding of the input sequence. For classification, the final hidden state corresponding to the special [CLS] token is isolated by the pooling layer. This single-vector representation, which aggregates information from the entire sequence, is then passed to the classification head. This head consists of two final components: (1) a fully connected logits layer that projects the high-dimensional feature vector into a 3-element vector of raw scores (logits), one for each target class; and (2) a softmax layer that normalizes these scores into a final probability distribution across the classes. For clarity, we refer to layers $\ell = 1, \ldots, 12$, whereas the machine learning literature convention is to number layers $\ell = 0, \ldots, 11$.

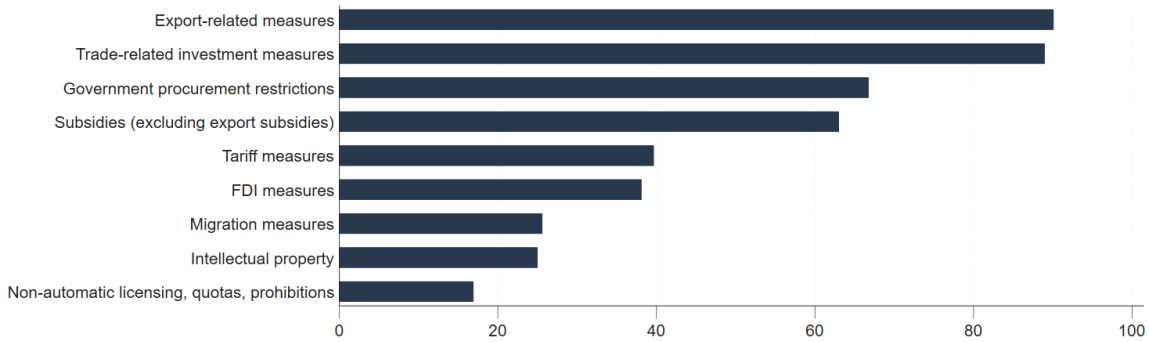


Figure G.2: Share of Policies Classified as Industrial Policy By Policy Instrument

Notes: This figure plots policies classified as industrial policy as a share of all policies with identifiable goals (i.e., excluding the not enough information class). Policy instruments defined based on UNCTAD's MAST chapter codes. We have added an additional category for import tariffs ("Tariff Measures").

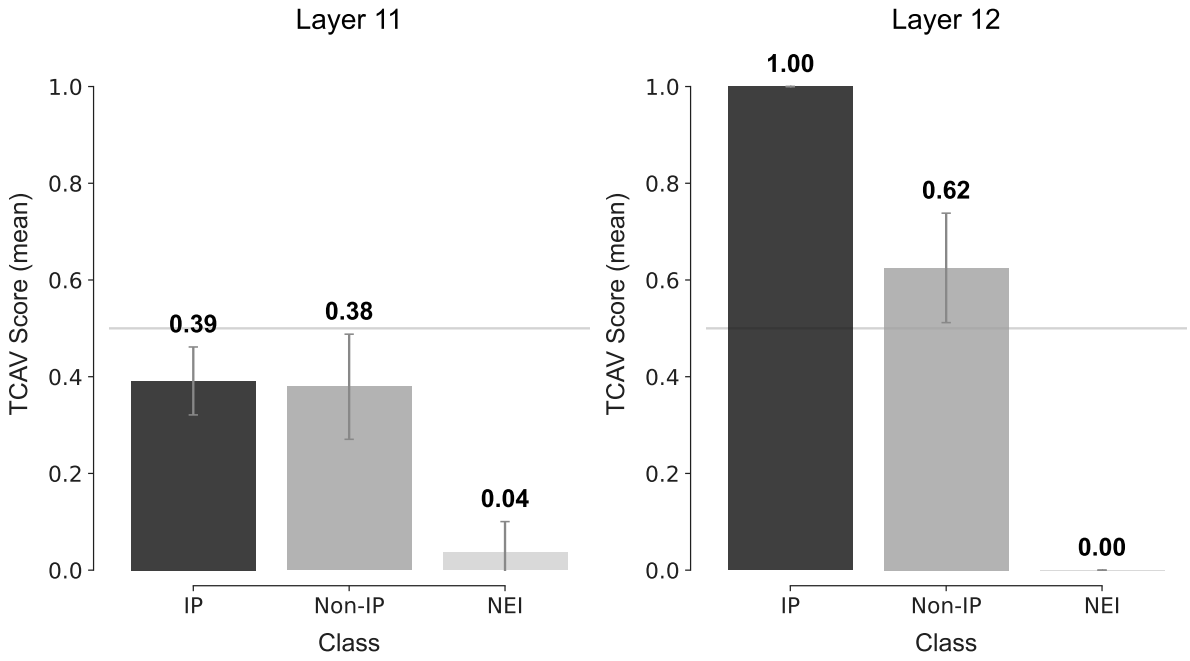
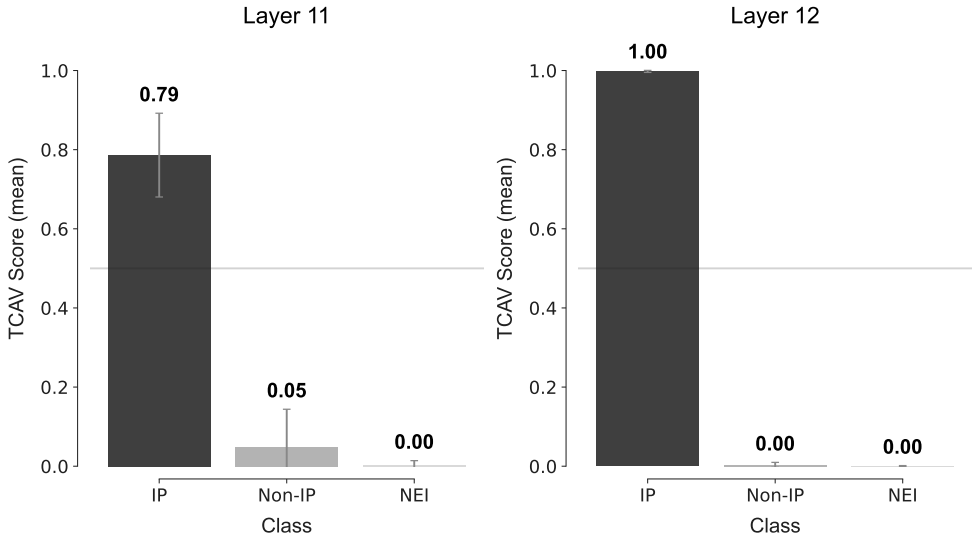
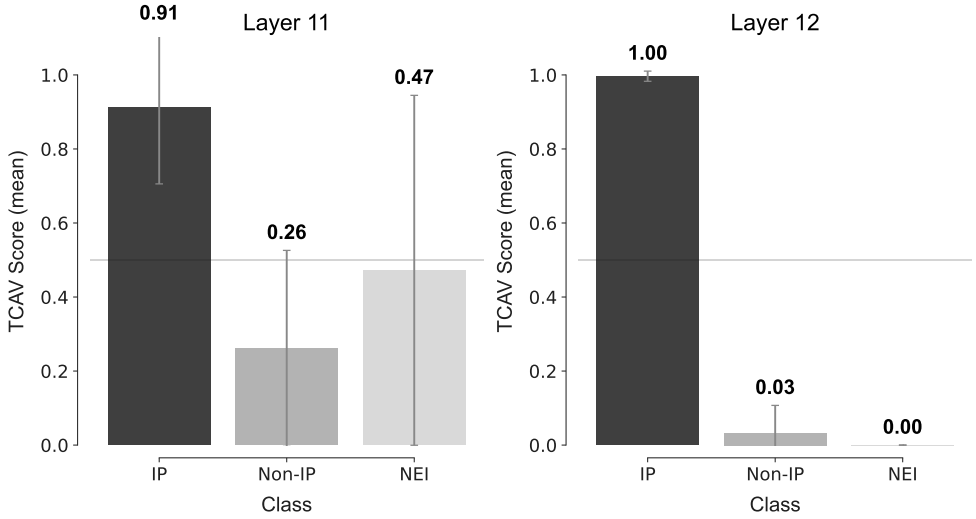


Figure G.3: Robustness: Average TCAV Scores for 'Industrial Policy Goals' Across Deep Layers, by Class (Alternative Mean Pooling)

Notes: This figure shows TCAV results for our fine-tuned BERT model using an alternative attribution method (mean pooling) for calculations. The analysis targets BERT's final two layers (11–12), which capture high-level semantic abstractions. Bars show TCAV scores for each class: "Industrial Policy", "Not Industrial Policy", and "Not Enough Information" (NEI). The horizontal line at 0.5 denotes TCAV scores that are as good as random. Scores were computed for 16 distinct negative concept sets (C versus $N_1, \ldots, N_{16}$). The y-axis reports the average TCAV score, and error bars show the standard deviation across the 16 comparisons. The TCAV score for Non-IP (.62) is larger than the baseline estimates and marginally above random. Robustness of Layer 12 scores suggest the results reflect the importance of the concept for industrial policy predictions.



(a) Default CLS Pooling



(b) Mean Pooling

Figure G.4: Robustness: TCAV and Mean TCAV Scores for 'Industrial Policy Goals' Across Deep Layers (Out-of-Distribution Negative Sets)

Notes: This figure shows TCAV results for our fine-tuned BERT model using alternative negative sets, $\mathcal{N}'_1, \ldots, \mathcal{N}'_{16}$, drawn randomly from out-of-distribution external data. Compared to in-distribution draws, Layer 12 TCAV scores remain similar, while Layer 11 scores are substantially larger. The latter means that the concept shows up earlier in the neural network. Robustness of Layer 12 scores suggest the results reflect the importance of the concept for industrial policy predictions.

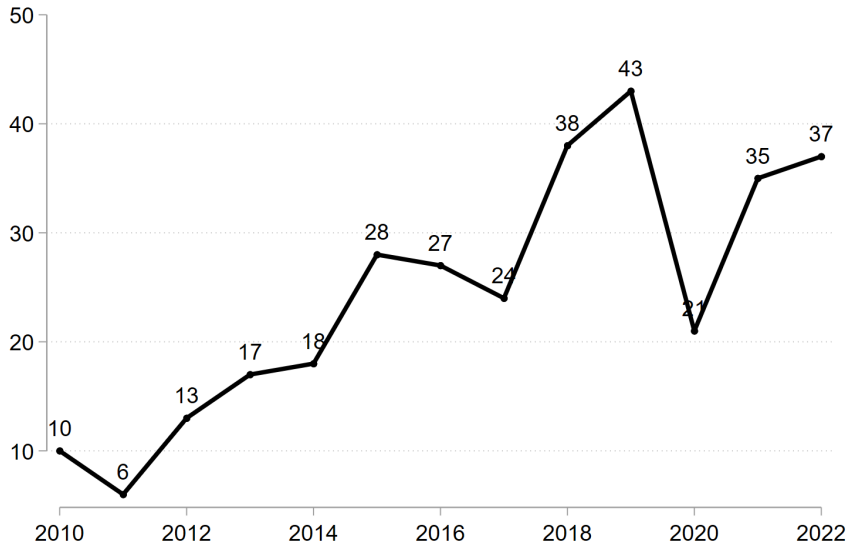
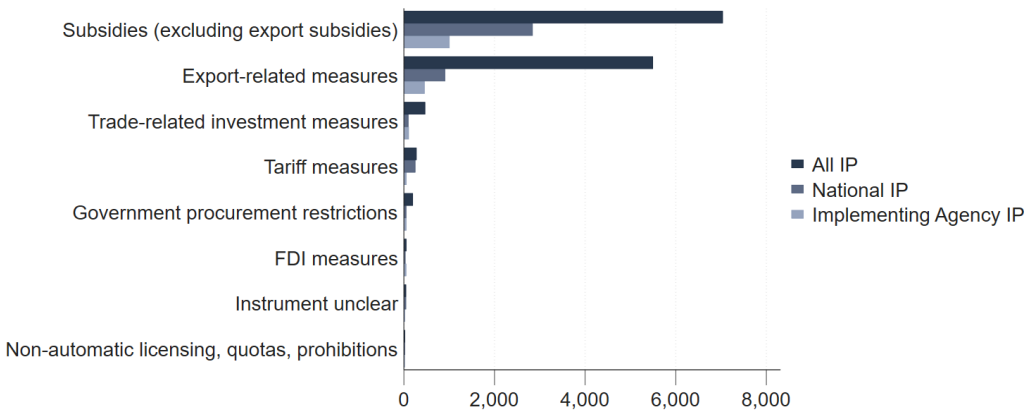
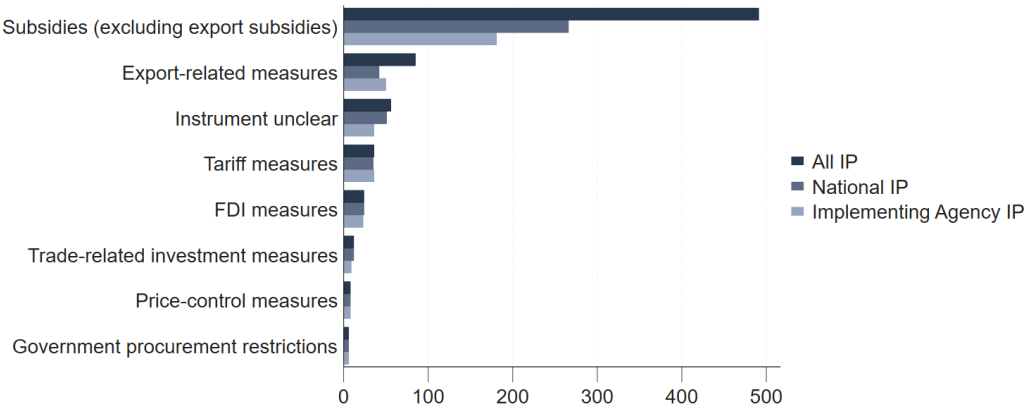


Figure G.5: Share of Industrial Policies (%) over Time

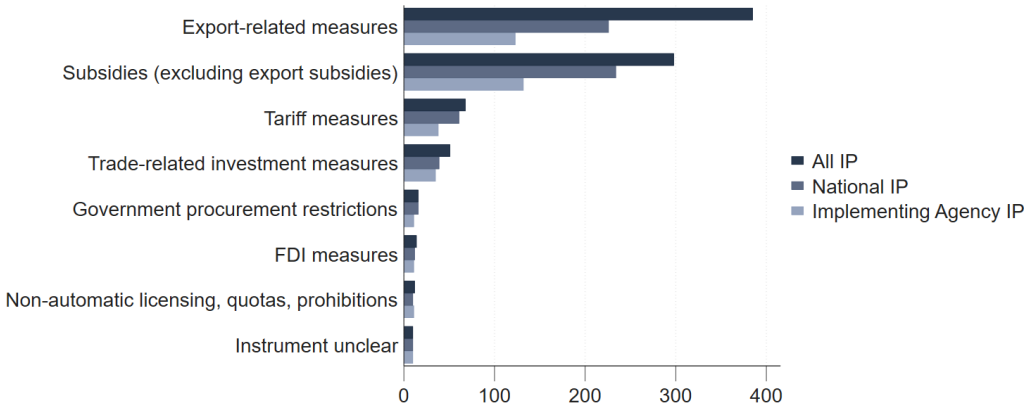
Notes: We calculate the share of industrial policies (%) out of all policies including those with and without identifiable goals. We follow GTA guidance and use only policies recorded by the GTA in the same year that they were announced for this exercise. This is due to the substantial back-filling of data in the GTA, which is a living dataset. By using only policies recorded in the same years as they were announced, we ensure the comparability of data across both more distant and recent years.



(a) High-Income



(b) Middle-Income



(c) Low-Income

Figure G.6: The Instruments of Industrial Policy by Income Group

Notes: For this figure, low, middle and high-income countries are those in quintiles 1 & 2, 3, and 4 & 5, respectively.

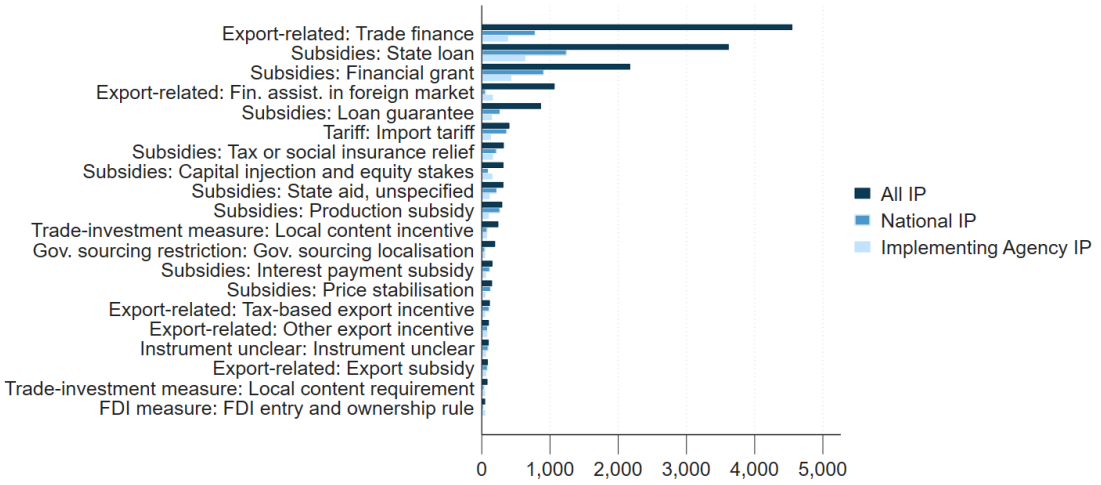
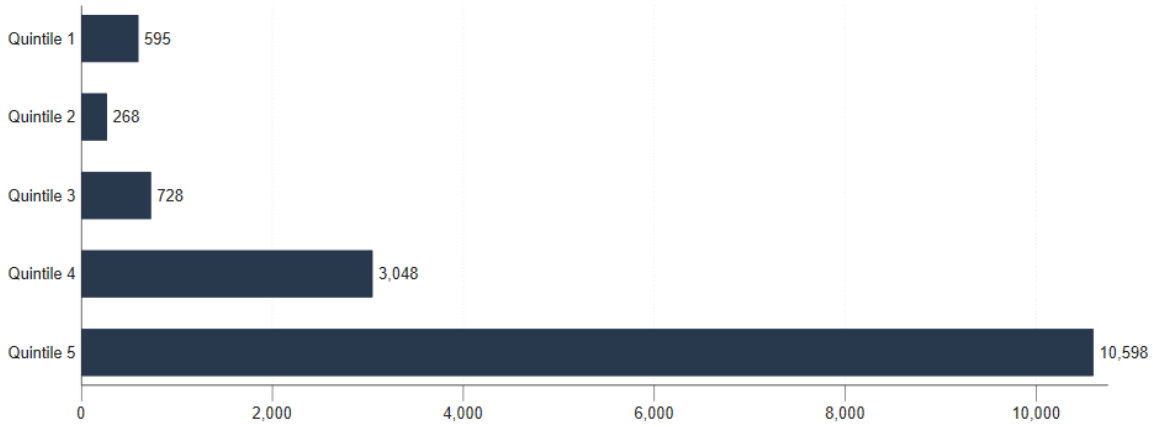
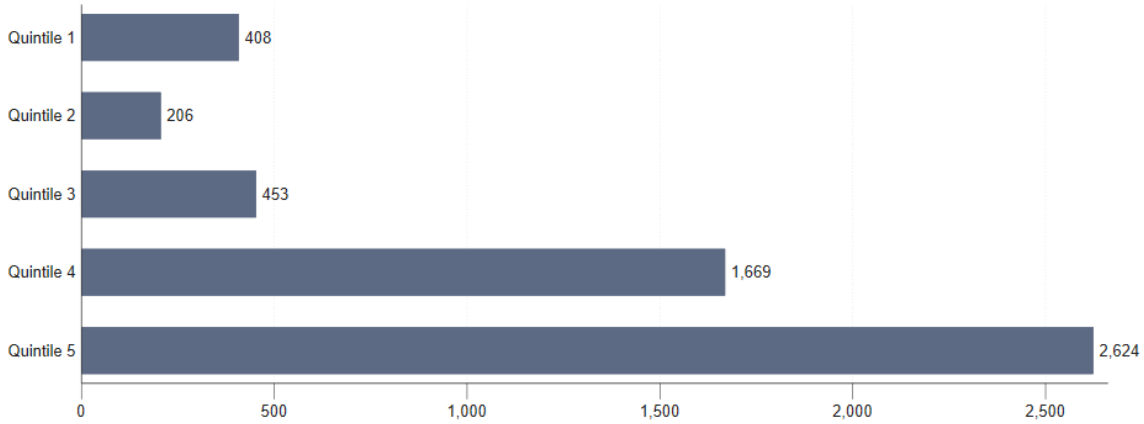


Figure G.7: The Instruments of Industrial Policy using GTA's In-House Classification System

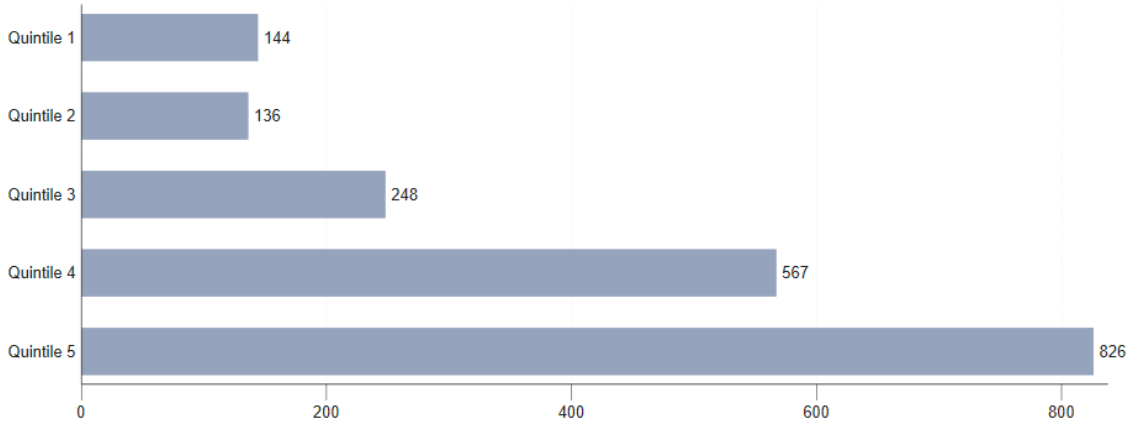
Notes: We present the twenty policy instruments that have the highest average usage ranking across the three measures of IP activity. The policy instruments excluded from the graph are: export bans, FDI financial incentives, export taxes, in-kind grants, local value-added incentives, state aid (nes), import tariff quotas, internal taxation of imports, import bans, import-related non-tariff measures (nes), export licensing requirements, labour market access restrictions, export quotas, public procurement preference margins, local labour requirements, import licensing requirements, import incentives, export-related non-tariff measures (nes), FDI treatment and operations (nes), local operations requirements, controls on commercial transactions and investment instruments, public procurement access restrictions, public procurement measures (nes), import quotas, controls on credit operations, local supply requirements for exports, export tariff quotas, post-migration treatment policies, intellectual property protection measures, local operations incentives, local value-added requirements, localisation measures (nes), trade payment measures, anti-dumping measures, and repatriation and surrender requirements. For Implementing Agency IP, we calculate the number of agencies implementing at least one industrial policy via each policy instrument in each year, and then sum these counts across years.



(a) All Industrial Policies



(b) National Industrial Policies



(c) Agencies Implementing Industrial Policy

Figure G.8: Industrial Policy Activity by Income Quintile

Notes: Panel (a) presents the simple aggregate sum of all the policies classified as industrial policies by our BERT model. Panel (b) excludes industrial policies directed at specific firms. For Panel (c), we calculate the number of agencies implementing at least one industrial policy in a given year and then sum these counts across years.

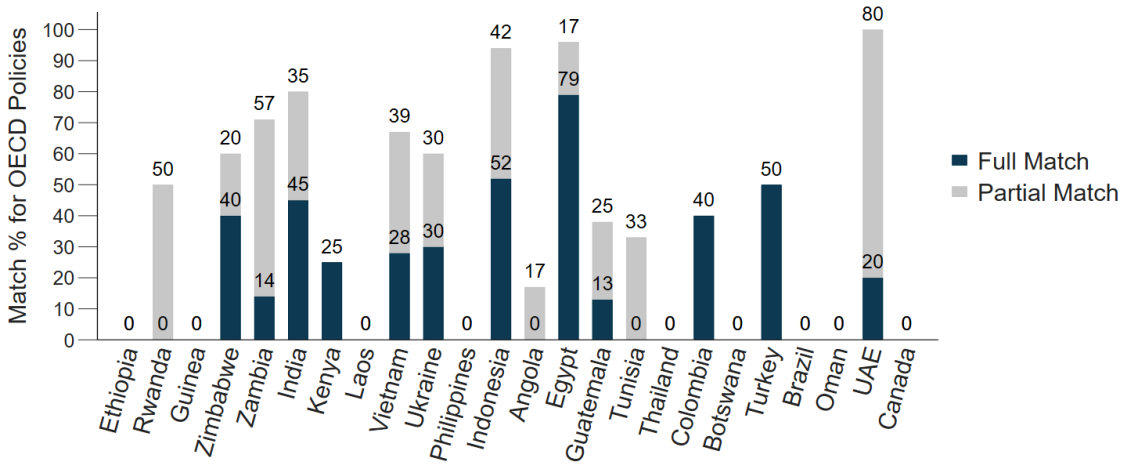


Figure G.9: Percentage of OECD Policies with GTA Matches by Country

Notes: Countries ordered from lowest (Ethiopia) to highest (Canada) 2010 GDP per capita. An OECD policy has a full match with a GTA policy if we can pinpoint the same policy document in the GTA. An OECD policy has a partial match with a GTA policy that uses the same 1) policy instrument, 2) affects the same industrial primary commodities, and 3) is announced within one year of the OECD policy being introduced. We provide the match rate out of all OECD policies in the GTA domain for each country. See Appendix D for more information.

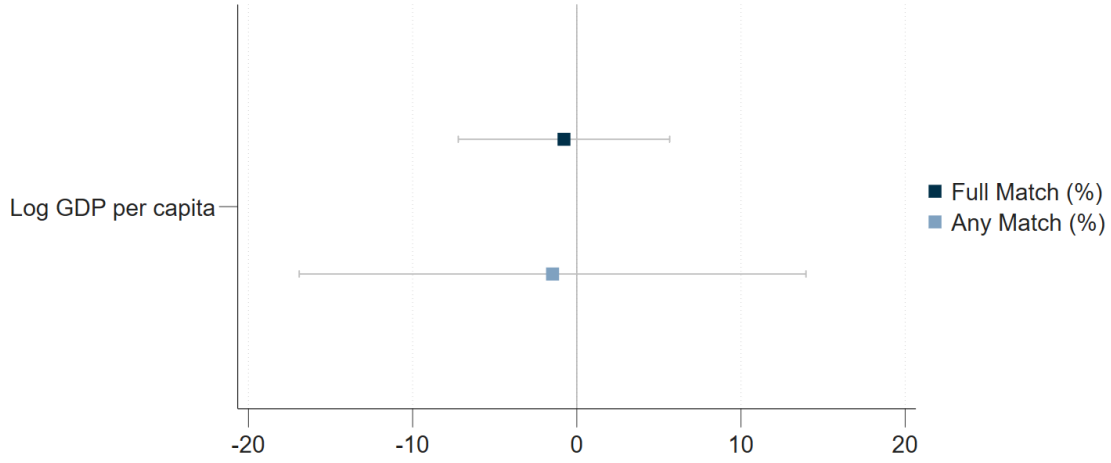


Figure G.10: Regression of Full and Any Match (%) on Log GDP per Capita

Notes: We regress the match rate (full or partial) on log GDP per capita and plot the coefficient of interest and its 95% confidence interval. The match rate refers to the share of policies in the OECD data that can be matched to a policy in the GTA. See Appendix D for more information.

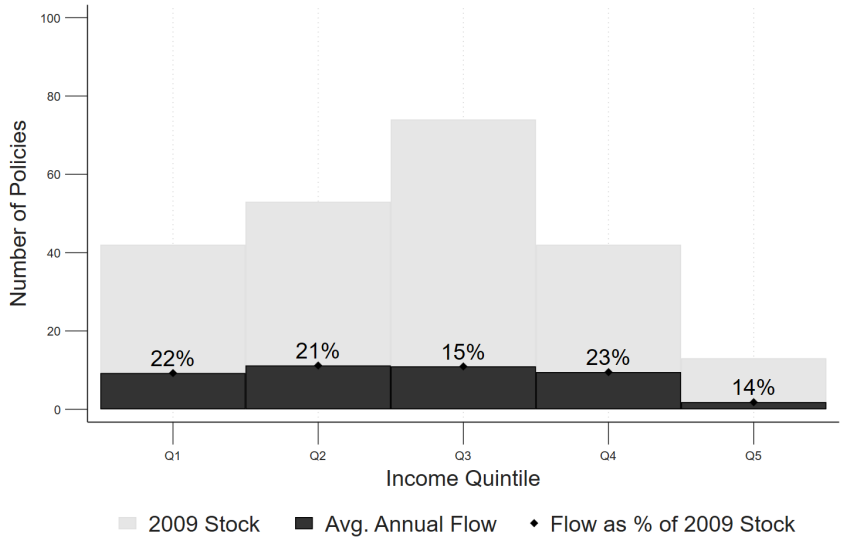


Figure G.11: Policy Flows vs 2009 Policy Stock by Income Quintile

Notes: Policy flows include all new or changing policies as proxied using the variable "Direction of Change" provided by OECD (2024). Quintiles based on 2010 GDP per capita. See Appendix D for more information.

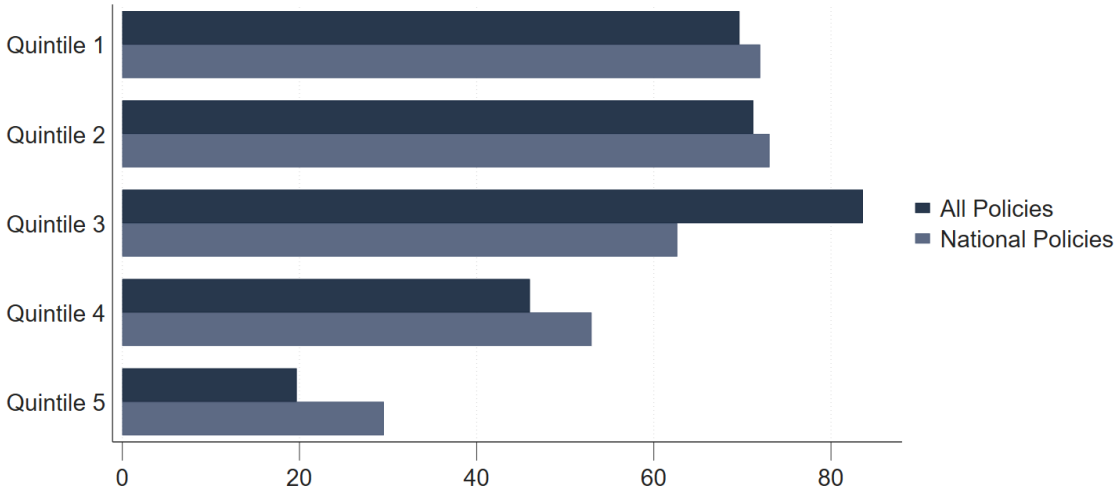


Figure G.12: Percentage of Policies Classified as "Not Enough Information" by BERT Three Class Model

Notes: This figure plots the share of policies classified as "Not Enough Information" by the BERT three-class model. "All Policies" is the baseline measure which enumerates all policies, "National policies" excludes those targeted at specific firms.

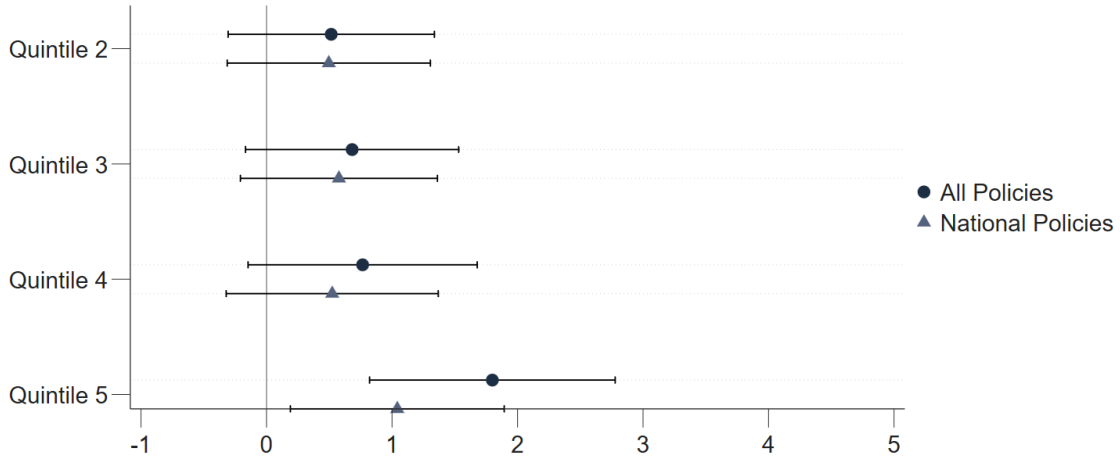


Figure G.13: Regression of Industrial Policy Activity Relabeling "Not Enough Information" Policies as Industrial Policies for Quintiles 1-3

Notes: We relabel all policies in quintiles 1-3 classified as "Not Enough Information" as "Industrial Policy" with the exception of two categories of policies we are confident do not contain industrial policies. These are 1) policies that have a duration of less than one month and 2) policies sourced from the WTO download facility which do not capture one policy, but rather all of the MFN, GSP or LDC tariff changes recorded by the WTO in that year. We then regress the log of measures of IP activity on income quintiles with the first quintile being the excluded category.

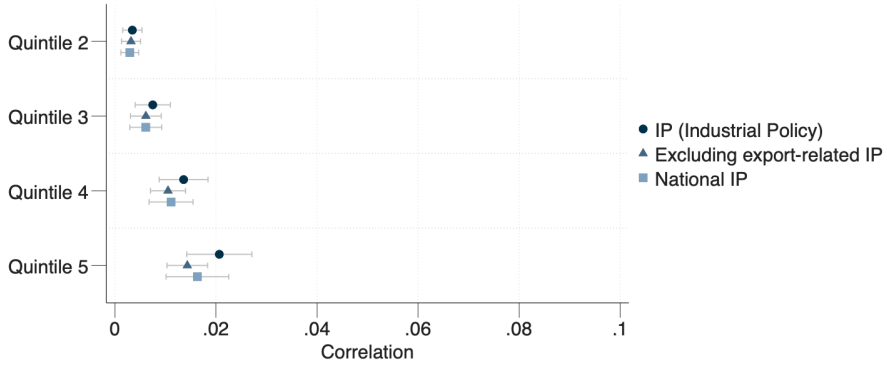


Figure G.14: Industrial Policy and RCA in High-Income Countries at the HS6 Level

Notes: This figure plots the estimated β_i coefficients from equation 3 and their corresponding 95% confidence intervals. We estimate the regression using data at the HS6 level. The dependent variables are as follows: "IP (Industrial Policy)" takes the value of one if a country-HS6 sector-year receives at least one industrial policy; "Excluding export-related IP" excludes industrial policies deployed via export-related measures; "National IP" takes the value of one if a country-HS2 sector-year receives at least one national industrial policy. The omitted category is the lowest quintile of the RCA distribution. High income refers to quintiles 4 and 5. All regressions include country-year-HS2 fixed effects and cluster standard errors by country.

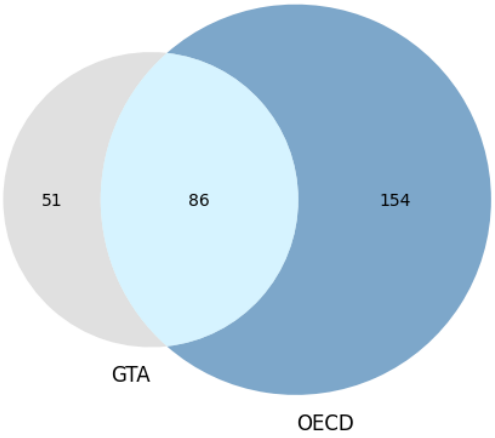


Figure G.15: Venn-Diagram of GTA and OECD Policies

Notes: Each unit in the Venn diagram is a unique piece of legislation or a policy document that is in the domain of the OECD and that is likely in the domain of the GTA. See Appendix D for more information.

H. Appendix Tables

Table H.1: Implementation Level for Policies

National	33011
Natonal Financial Institution	8764
International Financial Institution	3261
Supranational	2247
Subnational	1371
Total	48654

Notes: Distribution of the implementation level of the policies in the Global Trade Alert Database. We exclude the 1,371 subnational policies from our analysis.

Table H.2: Summary Statistics

	BERT Three-Class Model Prediction			
	Industrial Policy	**Other Intention**	**Not Enough Info.**	**All**
Panel (A): Implementation Level				
International Financial Institution	2204	742	315	3261
National Financial Institution	7269	404	1091	8764
National	5831	8179	19001	33011
Supranational	431	489	1327	2247
Panel (B): Firm-specific policies				
Not firm specific	5596	7133	14435	27164
Firm specific	10139	2681	7299	20119
Panel (C): MAST Chapter Code				
Capital control measures	15	200	215	430
Contingent trade-protective measures	0	2673	47	2720
Export-related measures	6004	660	2315	8979
FDI measures	93	151	782	1026
Finance measures	0	18	36	54
Government procurement restrictions	221	110	863	1194
Instrument unclear	115	87	324	526
Intellectual property	1	3	7	11
Migration measures	21	61	383	465
Non-automatic licensing, quotas, prohibitions	45	221	1779	2045
Pre-shipment inspection and other formalities	0	6	11	17
Price-control measures	18	95	402	515
Subsidies (excluding export subsidies)	8266	4850	7196	20312
Tariff measures	403	613	6436	7452
Technical barriers to trade	0	0	4	4
Trade-related investment measures	533	66	934	1533
Total	15735	9814	21734	47283

Notes: This table presents the distribution of the 47,283 interventions at the core of our analysis. We report the implementation level, whether the intervention is targeted at specific firms or has broader scope, and the corresponding MAST Chapter Code for each intervention, categorized according to the labels generated by the three-class BERT model.

Table H.3: Twenty Concept Examples for 'Industrial Policy Goals' (Randomly Sampled)

No.	Target Concept Sample
1	The program prioritizes rural infrastructure development to enhance agricultural production capacity.
2	Supporting the transformation of traditional retail into modern omnichannel services.
3	Designed to enhance capabilities in data analytics and business intelligence services.
4	With the goal of substantially increasing investment in and development of renewable energy infrastructure and generation capacity.
5	The strategy seeks to establish the nation as a leading hub for technology consulting and digital services.
6	With the purpose of creating horticultural research and development centers.
7	Develop national capabilities in edge computing and distributed processing technologies.
8	The strategy aims to develop competitive advantages in quantum computing and advanced computing technologies.
9	To enhance intellectual property safeguards to support technology and innovation trade.
10	To strategically expand and develop international markets for domestically produced goods and services.
11	Promoting the establishment of regional headquarters for multinational service corporations.
12	Targeting the creation of temperature-controlled supply chain systems for fresh agricultural products.
13	To establish special economic zones focused on export-oriented production.
14	To strengthen fresh produce supply chains from farm to consumer markets.
15	The initiative focuses on developing world-class educational services for international students.
16	Develop comprehensive digital literacy programs for workforce transformation.
17	Intended to promote growth in professional translation and localization services.
18	The framework prioritizes developing competencies in machine learning consulting and AI implementation services.
19	Seeking to build world-class capabilities in artificial intelligence and machine learning applications.
20	Prioritize onshore development of critical technologies.

Notes: This table presents 20 random examples from the 300 example concept set C. Concepts are generated using the definition of industrial policy objectives and generated across sector (services, manufacturing, agriculture) and activities.

Table H.4: Examples of Agencies Implementing Industrial Policy

Implementing Agency name	**Country**	**Policy Description (from GTA)**
Bangladesh Bank	Bangladesh	[...] *Bangladesh Bank*, while announcing the export subsidy for the year 2021-2022, added 4 categories to the list of goods that will be eligible for such a subsidy.
Banco Nacional de Desenvolvimento Econômico e Social	Brazil	[...] the *National Bank for Economic and Social Development* (BNDES in Portuguese) provided a BRL 77.83 million [...] loan to an undisclosed agricultural undertaking. The loan was issued under the PRONAF Investimento Faxia 2 scheme.
Government of Congo	Congo	[...] the *government of Congo* implemented an export ban on wood logs. According to news reports, this decision follows a governmental strategy to strengthen the wood processing industry within the country.
Société d'Enrichissement du Tricastin	France	[...] the European Investment Bank (EIB) and *Societe Denrichissement Du Tricastin (S.e.t)* signed an agreement worth EUR 400 million [...] for the project Areva Uranium Enrichment Facility from France.
Government of Ireland	Ireland	[. . .] *Ireland* introduced a EUR 1176 (USD 1507) million price stabilisation scheme. The scheme will support undertakings producing electricity through renewable energy sources.
Secretariat of Agriculture and Rural Development	Mexico	[...] the *Secretary of Agriculture and Rural Development of Mexico* published an Agreement with the operating rules of the new Agriculture, Livestock, Fisheries and Agriculture Promotion Programme for the 2022 fiscal year.
Government of Puerto Rico	Puerto Rico	[...] the *government of Puerto Rico* approved Act No. 20 otherwise known as the "Act to Promote the Export of Services".
Ministry of Public Enterprises	South Africa	[...] the *South African Ministry of Public Enterprises* announced allocating R10.5 billion (USD 640.7 million) to finalize the business and rescue plan of South African Airways (SAA).
Türkiye Cumhuriyet Merkez Bankasi	Turkey	The Decree 2013/4 of *the Money-Credit and Coordination Council* replaces the Decree 2010/10 and encompasses an export subsidy programme for agricultural products.
Federal Crop Insurance Corporation	United States of America	[...] the *Department of Agriculture* approved a production subsidy worth USD 13 million to multiple US agricultural producers. [...] The program is administered by the *Federal Crop Insurance Corporation (FCIC* under [the] *Risk Management Agency (RMA) of* USDA.

Notes: We provide a random example of a policy text associated with each implementing agency in a random sample. Policy descriptions (excerpts) from the Global Trade Alert. The implementing agency names were extracted from the policy text by us. There is not necessarily only one implementing agency identified per policy text. The text that refers to the names of the implementing agency/agencies associated with the policy text have been italicized by us.

Table H.5: Regression of IP Activity on Income Quintiles

	All Policies				National Policies				Agencies Implementing Policies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Quintile 2	0.35883	0.58394**	-0.27456	0.14065	0.31841	0.51698**	-0.23554	0.12641	0.30910	0.47134***	-0.12310
	(0.30681)	(0.22753)	(0.31655)	(0.17715)	(0.28443)	(0.20976)	(0.29165)	(0.16492)	(0.24177)	(0.17950)	(0.24390)
Quintile 3	0.69632**	1.60125***	-0.06286	0.32358*	0.61682*	1.41505***	-0.04726	0.35182**	0.52223**	1.17443***	0.00210
	(0.34584)	(0.28964)	(0.33460)	(0.17624)	(0.31294)	(0.25342)	(0.30331)	(0.16359)	(0.26107)	(0.21228)	(0.25068)
Quintile 4	1.84612***	2.71531***	0.77672**	0.78576***	1.62418***	2.39088***	0.68689*	0.77055***	1.24028***	1.86672***	0.50151*
	(0.42657)	(0.29697)	(0.38844)	(0.19712)	(0.38575)	(0.26336)	(0.35295)	(0.18754)	(0.30702)	(0.20506)	(0.28375)
Quintile 5	2.87985***	3.64386***	1.83121***	1.48968***	2.14277***	2.81670***	1.18951***	1.26446***	1.62018***	2.17082***	0.86065***
	(0.46242)	(0.28109)	(0.42279)	(0.16404)	(0.38975)	(0.22574)	(0.36613)	(0.17084)	(0.30570)	(0.18236)	(0.28997)
Log 2010 Population		0.69802***				0.61572***				0.50308***	
		(0.05117)				(0.04464)				(0.03599)	
Log Num. HS6 Traded			0.90937***				0.79717***				0.62743***
			(0.13828)				(0.12044)				(0.09495)
Log Total Policies				0.97921***							
				(0.03535)							
Log Total National Policies								0.90063***			
								(0.03702)			
Observations	185	185	176	185	185	185	176	185	185	185	176

Notes: We regress the log of measures of IP activity on income quintiles with the first quintile being the excluded category. Robust standard errors in parentheses. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels, respectively.

Table H.6: Performance of the Main Model Versus Models Fit on Income Groups

		Precision		Recall		F-1 Score		Support	
Sample	Class	Main Model	HI Model	Main Model	HI Model	Main Model	HI Model	Main Model	HI Model
Full Sample	IP Goal	0.913	0.728	0.913	0.829	0.913	0.775	104	210
	No IP Goal	0.959	0.893	0.934	0.878	0.947	0.885	76	295
	Not Enough Information	0.947	0.954	0.954	0.936	0.95	0.945	260	1225
Low-Middle Income	IP Goal	0.857	0.621	0.947	0.774	0.900	0.689	19	106
	No IP Goal	0.889	0.827	0.889	0.821	0.889	0.824	18	140
	Not Enough Information	0.977	0.954	0.963	0.927	0.970	0.940	135	859
High Income	IP Goal	0.925	0.865	0.914	0.906	0.919	0.885	81	85
	No IP Goal	1.000	0.939	0.946	0.816	0.972	0.873	37	38
	Not Enough Information	0.903	0.868	0.933	0.878	0.918	0.873	90	90

Notes: We compare two BERT models for predicting Industrial Policy: the Main Model, trained on a combined sample of observations from both Low-Middle Income and High-Income countries, and the High-Income (HI) Model, trained and validated exclusively on data from High-Income countries. Each model is evaluated on its own test set. In both cases, the full test set—which includes all available observations—is further decomposed into Low-Middle Income and High-Income subsets. Note that the subgroups may not sum exactly to the full test set, as some observations have missing income information or represent aggregated entities (e.g., "U.S. Mexico"). For the HI Model, the full sample test set includes all observations from Low-Middle Income countries, all observations without income group labels, and the remaining High-Income policies (those not used in training and validation). The Main Model includes observations from the full income distribution across its training, validation, and test sets. The table reports standard performance metrics (Precision, Recall, and F1 Score) for each model across three categories—"IP Goal", "No IP Goal", and "Not Enough Information"—first for the full test set and then separately for the Low-Middle Income and High-Income subsets. This comparison allows us to evaluate how well each model generalizes across different income contexts.

Table H.7: Comparison Model Performance on Low-Middle Income Common Test Set

	Precision		Recall		F-1 Score		Support	
Class	Main Model	HI Model	Main Model	HI Model	Main Model	HI Model	Main Model	HI Model
IP Goal	0.857	0.680	0.947	0.895	0.900	0.773	19	19
No IP Goal	0.889	0.800	0.889	0.889	0.889	0.842	18	18
Not Enough Information	0.977	0.969	0.963	0.911	0.970	0.939	135	135

Notes: We compare two BERT models for predicting Industrial Policy: the Main Model, trained on a combined sample of observations from both Low-Middle Income and High-Income countries, and the High-Income (HI) Model, trained and validated exclusively on data from High-Income countries. Each model is evaluated on the same test set of observations from Low-Middle Income countries' policies. In both models these are out of sample observations, not used in training or validation. The table reports standard performance metrics (Precision, Recall, and F1 Score) for each model across three categories—"IP Goal", "No IP Goal", and "Not Enough Information". This comparison allows us to evaluate how well each model generalizes across different income contexts.

Table H.8: Predictive Performance of Fine-Tuned Bert Model on Annotated Splits

Data Split	Class/Metrics	Precision	Recall	F1-Score	Support
Train	IP Goal	0.980	0.993	0.987	448
	No IP Goal	0.988	0.994	0.991	329
	Not Enough Information	0.996	0.989	0.993	1128
	Macro Avg	0.988	0.992	0.990	1905
	Weighted Avg	0.991	0.991	0.991	1905
Test	IP Goal	0.913	0.913	0.913	104
	No IP Goal	0.959	0.934	0.947	76
	Not Enough Information	0.947	0.954	0.950	260
	Macro Avg	0.940	0.934	0.937	440
	Weighted Avg	0.941	0.941	0.941	440
Validation	IP Goal	0.971	0.978	0.975	138
	No IP Goal	1.000	0.990	0.995	102
	Not Enough Information	0.991	0.991	0.991	347
	Macro Avg	0.988	0.987	0.987	587
	Weighted Avg	0.988	0.988	0.988	587

Notes: This table presents the predictive performance of the BERT model across different data splits: train, test, and validation. For each split, it details standard classification metrics—Precision, Recall, and F1-Score—for the three class BERT Model. The classes are 'IP Goal', 'No IP Goal', and 'Not Enough Information'. Additionally, it provides the Macro Average and Weighted Average for these metrics across classes. The 'Support' column indicates the number of observations for each class (or average) by sample split.

I. Codebook Appendix

I.1. Codebook for identifying industrial policy intention from policy descriptions

You will be annotating, or coding, descriptions of economic policy. These policy descriptions you will code come from our Global Trade Alert (GTA) database. The following codebook introduces annotators (you) to the definitions and criteria used to code the intentionality of policy based on the measures description. Specifically, you will be coding whether or not policies show industrial policy intentions.

We take you through the coding process in four steps. First, we provide our working definition of "intentionality" as applied to industrial policy. Second, we describe the three different types of intentionality you will code. Third, we then provide a guide to annotation using Prodigy–our interface for coding policy text. Last, we give a series of examples of annotations; each example provides a detailed description of how and why we coded the examples.

I.2. Definitions

1. *Industrial Policy Intentionality.* Let us start with the definition of industrial policy "intentionality," in light of the text you will be annotating. Formally,

Definition 1 (Industrial Policy Intentionality)**.** *A policy or measure has an Industrial Policy Intentionality when it (i) seeks to change the relative prices across sectors or direct resources towards certain selectively targeted activities (e.g., exporting, R&D), with (ii) the purpose of shifting the long-run composition of economic activity.*

In other words, a policy or measure has Industrial Policy Intentionality when it is used for industrial policy goals. These policies have industrial policy purposes as opposed to other purposes: health, sanitation, national security, retaliatory measures, anti-dumping, safeguard measures, general SME (and also "midcap") entrepreneurship, or attempts to boost aggregate employment, etc.

This intentionality (Definition 1) should be clearly communicated and discerned from the text.

I.3. Three Types of Intentionality

As an annotator, your goal is to code policy into three categories of intentionality: 1) IP intention, 2) other (or non-IP) intention, and 3) not enough information. Using definition 1, every policy description can be classified as these three forms of intentional policy. More precisely, the three categories are,

1. Industrial Policy Intentionality ("IP intention") - A policy description is defined as having an industrial policy intention if it makes an explicit mention of an industrial policy objective, per the definition in Section 1.

2. Intention other than Industrial Policy ("Other intention") - A policy description is defined as having an intention other than Industrial Policy if it gives a proximate cause for the policy other than an industrial development aim. An incomplete list of examples: health, sanitation, migration and labor bans, currency stability, national security, retaliatory measures, antidumping, or safeguard measures, general SME entrepreneurship, or attempts to boost aggregate employment, etc.

3. Not enough information to discern intention ("Not enough information") - Descriptions without a clear intention, aim, or objective are included in this category. These are entries that describe a policy without stating its objective or rationale. Thus, there is not enough information to classify its intention.

I.4. Rules for Practice - How to Code Intentionality

How do we put the definition in Part 1 into action? Below are important, practical points for applying our definition to the data whilst coding. Use these lists while you code, and refer to them alongside the flow chart at the end of this section shown in figure I.2.

1. Basics Rules.

- **Look for "the why":** Search for "the why" in the policy descriptions. Look for the policy's own reasons. Some policies will state their reasons. Others will have implicit reasons (e.g. policies that are "sanitary restrictions" or a policy named "The Programme to Promote Growth in The Noodles Industry").

- **It is okay to assign cases to Type 3-"not enough information available.":** There may be many cases where there is not enough information to determine intention.

- **Take policies at face value:** Use the information in the policy description to make your coding decision. Take their goals at face value. Minimize using external knowledge to inform your assessments (e.g., "technical criteria are frequently used as de facto protectionism" cannot affect an annotator's reading of the intention behind any technical criteria).

- **Work independently:** It is crucial that each annotator work independently. If you are unsure about how to classify a measure, make your best judgment. Do not discuss these answers with other annotators. Discrepancies across annotators are important information for us to retain. If big questions arise, feel free to ask us.

- **Reasonable people may disagree - there may not be a "right" answer:** Some policies will be clear-cut and easy to code. Other times, however, intentionality will not be easy to detect. Thus, reasonable people may disagree on whether there is sufficient evidence of intent. Follow the guidelines above and make your best judgment. Know that there may not be a single correct answer. A diversity of answers is useful.

2. Applying Definitions: Heuristics and Tips.

- **De facto effects are not intentions:** When it comes to identifying industrial policy intentions, we are after policies with clear goals. Some measures, however, will have the effect – that is, de facto effect – of changing the long-run composition of economic activity, even if this is not the policy's goal.
- **These de facto policies are not "intentional" policies**. For example, if the stated aim is national security alone, this is not evidence of IP intentionality, although national security policies may change the long-run composition of economic activity.
- **"Selectivity" is useful:** Appeal to selectivity. It will help to distinguish between IP and non-IP intentions. This is because policy descriptions may not provide much information to the coder (you). For example, is a state trying to boost aggregate employment, or is it trying to boost employment in selective ways (e.g. by fostering specific "good" jobs)? The former is not IP intent, while the latter is. In these cases, the idea of selectivity" from our definition (above, Definition 1) is helpful.
- **The "long-run" matters. Question short-run measures** Our definition (Part 1) uses the term "long-run." However, You do not need to identify language that explicitly states that a policy is "long-term." Rather, the "long-run" part of the definition rules out temporary government interventions for fluctuations and business cycle reasons.
- **Some policies, at face value, show industrial policy intentions:** These policies include export promotion and R&D promotion. These measures are examples where the intentionality is "implicit" (see "Look for the why", above). We take these cases to be intentional industrial policy based on the type of measure in and of itself.

    a) **Export promotion:** Policies that promote exports through i) export subsidies, ii) export financing (bank loans etc.) or iii) by providing funding to agencies that do i) or ii) are classified as having IP intentionality. All of these export policies (i-iii) are costly (i.e., they are not in place with the intention of raising revenue) and often worsen a country's terms of trade (ToT). It's unlikely a policymaker would use them for any other purpose than promoting exports–a selective activity. Be careful, however—this does not include export quotas, export duties or export tariffs, which are more complicated. These policies tend to raise revenue or create rents. For these cases, we need to see more explicit intentionality for their use, unlike the clear-cut export promotion activities above.

    b) **R&D promotion:** Policies for R&D are selectively targeted and qualify as IP intention by definition. Like export promotion, R&D is costly, and its goal is often tied to the government's promotion of such technological activity with social (non-private) aims.

    Be mindful of similar policies that may–by their name–signal industrial policy intent.

- **Be careful of local content requirements and preferential procurement.** Though these policies may be industrial policies, they tend to be less selective than other IP and they may have other intentions (e.g. boosting aggregate employment, or national security reasons). In contrast with R&D and export promotion measures above, local content measures should not automatically be coded as IP intention at face value, unless there is clear selectivity (see below) or an explicitly stated IP goal.

- **Policies with "multiple intents" are classified as IP intention.** Some policies may have multiple goals. If a policy has IP intention, as well as other, non-IP aims, classify it as IP. Similarly, if an entry has multiple policies, some of which exhibit IP intention, and some non-IP intention, classify it as IP. We show examples of this below (in Section 4).

We now turn to actually coding the content using our web interface.

I.5. Annotating Policies in Prodi.gy

We use Prodigy to annotate the policy descriptions,which come from the GTA database. Prodigy is a simple annotation interface for saving and tracking your progress. Each annotator will receive a personalized link to their own Prodigy interface. The interface allows you to code policy descriptions quickly and transparently. Upon opening your Prodigy link, you will see the screen in figure I.1. The body of text is the policy description from our GTA database. Some will be long, others will be extremely concise. Beneath the policy description are three checkboxes. Each box corresponds to one of the three types of "intentionality" described in Section 2.

Select one of the three boxes that best describes the policy description, given the definitions we have provided. Choose only one. To make things as clear as possible, we break down the steps for annotating below.

1. Read the policy description thoroughly

2. Choose one of the three boxes that best describes the intentionality of the policy. That is, select the cell in the annotation area (image 2) that contains the best corresponding category. Note that once you pick a box, prodi.gy will automatically "accept" the annotation and move to the next example. You can always go back to the previous example by choosing it in the "History' section on the bottom-left corner. Note however, that once you "save" an annotation (with the diskette icon on the top left in image 2), you cannot make changes or view the policy again.

3. After labeling the policy description, follow one of the next options to continue annotating:

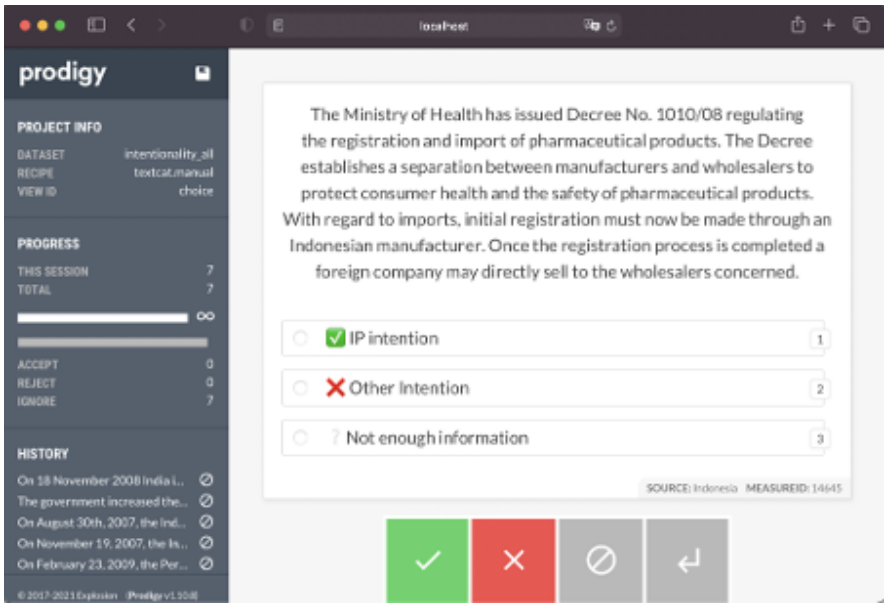


Figure I.1: Prodi.gy interface

a) **Accept** - Prodigy will automatically accept your annotation once you select a box with a label

b) **Reject** - If there's a critical spelling error in the entry, or there are strange symbols preventing you from reading the text, or the entry is empty, please press the reject button at the bottom of the screen and reject the annotation so the model doesn't learn from it. This will allow us to review the entry and fix the error.

c) **Pass** - If you're unsure of your annotation, meaning you don't know which label fits better for the text, or to which category the entry belongs to, press the ignore button at the bottom of the screen to pass on that entry (the model won't learn from it).

   Note, that "Pass" is different from selecting the "not enough information label". The "not enough information" label should be chosen when you are sure that there is simply not enough information in the text to determine the intention of the policy measure. Only skip entries (i.e., press "Pass") when you're not sure which of the three labels apply to a policy description.

d) Continue annotating.

e) If you make a mistake you can go back to the previous annotation by pressing the button with the back arrow at the bottom of the interface. If it's an older annotation you can find it in the History section in the bottom left corner of the interface.

f) Save often. Every few minutes while annotating, save your annotations in the diskette at the top. When you save, will not be able to go back and re-edit annotations.

g) At the end, save your annotations before exiting your session. Be sure to click the "save" button once you're done.

I.6. Example Identifications and Their Logic

Annotation is not always straightforward. Trust us, we have read many of these. Thus, in this section we provide concrete examples for annotating. We cover some of the cases that fall into one of the three categories spelled out in Part 2 (section 1), as well as a logical decision tree template that you should use when evaluating each entry 2.

1. Lists of intentionality examples (including examples of "not enough information cases"). The three tables below provide examples of the three types of "intentionality" you will code. Each table corresponds to one of the three types of policies you will code; we show you the policy descriptions and the logic for their classification according to our definition.

Table I.1: IP Intentionality Examples

#	Description	Details
1	On 1 March 2013, Nigeria renewed its certification criteria for the import of goods and launched a new Conformity Assessment Program (SONCAP). The main inspections are now done before the shipping and will accelerate transportation. Furthermore, the Standards Organization of Nigeria (SON) will check if the goods are in conformity with respect to the Nigerian standards, which in many cases are international standards. Hence, the mutual recognition of quality standards will facilitate compliance and allow for smoother customs administration. Furthermore, the development will also facilitate the export of Nigerian products. Nevertheless, various goods are excepted from the list, namely: food products, drugs, chemicals used as raw materials, military equipment, aviation related products, CKD bicycle, motorcycle, automobiles and industrial machinery.	**Intentionality label:** IP intention. **Explanation:** This is a tricky entry as the true intent only becomes clearly IP towards the end: facilitate the export of Nigerian products

Continued on the next page

#	Description (Continued)	Details (Continued)
2	On the 20th of March 2017, the Chinese General Administration of Customs announced its first official implementation following Premier Li Keqiang's 'executive meeting' (see source 5) with the State Council in December of 2016, outwardly urging for more policies that favour foreign investment within China. The removes all taxes and tariffs on imports pertaining to foreign investment entreprises listed in an announcement from the National Development and Reform Council and Ministry of Commerce (2017 revision). The affected industries differ from province to province, focusing on areas that are already points of economic strength for each region. Broadly speaking, the following sectors are prioritised for investment: Agriculture technology and derived products, Mining, Infrastructure, Tourism, Traditional medicine. This brings current policy back into line with a 1997 State Council edict, which was the first to implement these tax and tariff exemptions for foreign investors. The 1997 policy was changed in 2008, exempting relevant firms only from import tariffs. VAT and other import taxes still applied, until now.	**Intentionality label:** IP intention. **Explanation:** Attraction of foreign FDI in areas of economic strength. The statement of intent is a little weak, but focusing FDI on areas of economic strength hints at an urge to promote these sectors. Notion of selectivity in that FDI is incentivised selectively in areas of economic strength and this varies province by province.
3	On the 8th of November 2016, the Provincial Government of Shaanxi, China, announced its regional implementation of the PRC Ministry of Industry and Information Technology's 13th Five-Year Plan for the textile industry. The goals of the regional plan are largely the same, with targets mentioned such as the promotion of green production technology, maintenance of the strength of the Chinese textile industry both at home and overseas, etc. In Shaanxi, however, there is an emphasis on the use of preferential financial treatment, as opposed to tax breaks and grants, in carrying out the plan. The plan mentions the improvement of credit support; supportive financial policies for new firms and the establishment of an 'Equity Trading Centre' to give bonuses to successful entreprises. The plan will apply for the length of the national 13th Five-Year Plan Period: 2016-2020.	**Intentionality label:** IP intention. **Explanation:** Implementation of a major industrial plan.
4	On 17 May 2011, the French government passed an amendment to its Language Law. The amendment sets a price floor for the sale of books. The floor is equal to 95% of the price stated by the French publishers. This amendment also extends to e-books. Thus, online book retailers have to sell ebooks at similar prices to the hard copy versions. According to the Spring 2014 Global eBook report by Ruediger Wischenbart (p. 36), France's online book selling market is dominated by the American companies Amazon, Apple and Google Books - with the exception of Fnac, France's largest book chain. Therefore, while this amendment does not appear to involve any de jure discrimination against foreign commercial interests, its implementation de facto discriminates against foreign firms that happen to be large players in the e-book market. Hence, the amber classification of this measure. The amendment came into force on 26 May 2011.	**Intentionality label:** IP intention. **Explanation:** Entry states that the "implementation [of the policy] de facto discriminates against foreign firms", thereby providing more favorable conditions to domestic producers.

Continued on the next page

#	Description (Continued)	Details (Continued)
5	The Clean Energy Finance Corporation financing is subject to the Australian Industry Participation (AIP) policy. This framework states that all programs enclosed in this policy must encourage the participation of Australian companies in major public and private projects carried in the country. In this sense, companies applying for a CEFC credit line must provide an AIP Plan to demonstrate the strategy to maximize opportunities for Australian industry to participate in the project. Therefore, it can be understood that the AUD 100 million (over USD 71.3 million) finance allocated to RateSetter is subject to local content requirements.	**Intentionality label:** IP intention. **Explanation:** Entry outlines that Australian Industrial Participation plans require companies to "maximize opportunities for Australian industry".
6	On July 19th 2010, the Brazilian government, through the 'Medida Provisória nş 495' introduced changes in Law Nş 8.666, which establishes the general rules regarding administrative contracts and governmental procurement related to works, services, including marketing, acquisitions, sales and leases carried out by the three levels of government. Among the main modifications in the Brazilian rules introduced by 'Medida Provisória Nş 495' is the establishment of a level of preference of up to 25% above the price of external manufactured goods or services, to be granted to local manufactured products, or national services or group of products or services, that comply with technical local regulations. The level of preference to be granted will be established according with criteria related to the creation of revenue and employment, the fiscal impact and national technological innovation.	**Intentionality label:** IP intention. **Explanation:** Level of preference related to creation of national technological innovation.
7	On April 16, 2018 Minister of Innovation, Science and Economic Development, Montréal, Quebec Navdeep Bains announced a Can.$49.5 million (U.S.$38.3 million) subsidy in an aerospace consortium led by Bell Helicopter Textron Canada Ltd. The funding will help Bell and 18 industry and academic partners develop technologies to be integrated into next-generation helicopters, which can fly with or without a crew on board, and fully autonomous aerial systems. Other innovations include technologies to make aircraft more energy efficient and environmentally sustainable as well as technology to reduce noise pollution. The 18 industry and academic partners include Pratt & Whitney Canada, CMC Electronics, Esterline Technologies Corporation, several small and medium-sized businesses, and nine Canadian universities.	**Intentionality label:** IP intention. **Explanation:** Development of a new technology.

Continued on the next page

#	Description (Continued)	Details (Continued)
8	The Industry Ministry has agreed to give Rp 50 billion (USD 5 million) to state owned sugar makers in a bid to revitalize the sugar industry and to attain self-sufficiency for all domestic demand, both for households and for industrial use, by 2014. The fund will be distributed in the form of 10 percent subsidy for every purchase of new machinery by nine state sugar companies (including PT Perkebunan Negara and subsidiaries, PT Rajawali Nusantara Indonesia and subsidiaries, and PT Madu Baru - a joint venture between the Yogyakarta Sultanate and the government). According to Director General for Metal, Machinery, Textiles and Miscellaneous, Ansari Bukhari, the subsidy does come with a condition being that machines purchased by the companies must be entirely assembled in Indonesia and with a minimum 40 percent local content (the GTA identified affected tariff line for sugar machinery is 8438). Under the scheme, the companies must first buy the new machines and then request reimbursement by the Industry Ministry, with validation by the Agriculture Ministry. As the main purpose of this program is to reach self-sufficiency in the sugar sector, the affected trading partners are identified as the exporters to Indonesia of more than USD one million in trade value for at least one of the identified tariff lines.	**Intentionality label:** IP intention. **Explanation:** Attempt to revitalize the sugar industry.
9	In March 2013, the Belarusian government allowed a second Russian bank to provide export financing for Belarusian machinery exports to Russia. With Decree 176 of March 13, 2013, the Council of Ministers of Belarus included VTB Bank in a state financial scheme. The objective of this measure is the provision of loans at advantageous terms to buyers of Belarusian machinery on the territory of the Russian Federation. Since 2009 this Belarusian state support scheme was only made available through Sberbank Russia. The favourable credit terms are guaranteed on the basis of Decree 466 of September 24, 2009 through partial reimbursement of interest payments by the Belarusian Ministry of Finance. An official press release of the Council of Ministers of Belarus explains that the goal of attracting a second bank (VTB Bank) in this state initiative is to promote the Belarusian goods on the Russian market and to increase the Belarusian exports. The GTA includes state guarantees and other financial incentives that are likely to affect the restructuring and performance of firms facing international competition, whether from imports, in export markets, and from foreign subsidiaries.	**Intentionality label:** IP intention. **Explanation:** Aims to promote Belarusian machinery exports in the Russian market.

Continued on the next page

#	Description (Continued)	Details (Continued)
10	On 12 March 2018, the Instituto de Crédito Oficial (ICO) signed a financing agreement with Spanish Acerinox having a total value of EUR 100 million (approx. USD 123.4 million). The loan will support the North American Stainless' project concerning establishing two new production lines in order to increase the production of bright finish stainless steel and expand the product range of the company's final goods. The two new lines are a: "BA-finish bright annealing line and a cold rolling mill". North American Stainless is an American subsidiary of the Spanish Acerinox Group. In this context, the CEO of Acerinox state in a press release: "this agreement will enable us to reaffirm our leadership in the US market and increase our competitiveness". Instituto de Crédito Oficial is a state-owned bank whose function is to promote "economic activities contributing to growth, the development of the country and improving the distribution of the national wealth." Among other activities, the bank manages Spain's official funding instrument to promote Spanish exports and development aid. A state act in the GTA database is assessed solely in terms of the extent to which its implementation affects foreign commercial interests. On this metric, the investment support granted here is discriminatory.	**Intentionality label:** IP intention. **Explanation:** Increase the competitiveness of an industry.
11	With Decree 511 of 14 November 2013, the President of Belarus approved the provision of an investment loan (EUR 8.53million) in 2014-2016 from the state budgetary fund for national development to the export-oriented state-owned textile enterprise JSC "Sukno". Among its main export products are: woollen and semi-woollen fabrics, blankets and plaids as well as defence materials. Furthermore, the government authorised the issuance of a state guarantee to cover the 2013-2016 (EUR 45.80million) and 2014-2015 (EUR 11.23million) loans, to be extended by Belarusbank. Their purpose is to increase the sales profitability and labour productivity in 2013-2024 of the enterprise. The first loan must be repaid not later than 31 December 2024; and the second in the period 2017-2023. However, the issuance and maturity dates of the state guarantee were not disclosed. The GTA includes state guarantees and other financial incentives that are likely to affect the restructuring and performance of firms facing international competition, whether from imports, in export markets, and from foreign subsidiaries.	**Intentionality label:** IP intention. **Explanation:** Increase the export performance of an industry.

Continued on the next page

#	Description (Continued)	Details (Continued)
12	On July 26, 2017 the governor of Michigan signed into law a set of three bills (SB242, SB243, and SB244) that are collectively referred to as the Good Jobs for Michigan Program. They create this program within the Michigan Strategic Fund (MSF) and a related fund within the Department of Treasury. Among other things, the program provides that authorized businesses may "capture" state income taxes withheld from certified new employees, subject to approval by the MSF, as an incentive to create new jobs in Michigan. This incentive is available both to businesses already operating in Michigan and those newly locating in the state. A business location or expansion project would require a resolution of approval from the local governing body. The share of taxes that a business could capture is a function of the number of jobs and the level of the wages that it creates. The MSF can enter into no more than 15 agreements each year, and cannot disburse more than $200 million in total withholding tax capture revenues over the life of the program. No new agreements can be entered into after December 31, 2019. Professional sports stadiums, casinos, retail businesses, and those portions of eligible businesses used exclusively for retail sales are not eligible.	**Intentionality label:** IP intention. **Explanation:** The objective is to create "good jobs", with a wage criteria. Selectivity helps to classify this example. Note that this has more selectivity than an aim pursuing aggregate employment (which we classify as not IP). The state is trying to change the composition of economic activity by creating incentives for "good" (i.e. high-paying jobs).
13	The National Union of Agricultural Insurers (according to Rossiyskaya Gazeta, an official newspaper of the Russian Government, Issue 6230 of 12 November 2013) has developed new state-supported insurance products for the Russian agricultural producers. This state measure is in line with the statement of the Russian President, Mr. Vladimir Putin (according to Rossiyskaya Gazeta, Issue 6187 of 20 September 2013), that agro producers must be backed up with stable guarantees and compensated in case of crop loss and other incurred risks.	**Intentionality label:** IP intention. **Explanation:** Promotion of agriculture.
14	On January 8, 2009 the government of Egypt eliminated a 2% export tax on Egyptian-made cars and exempted component parts from import tax. These measures were taken in order to boost domestic industry by making Egyptian cars more competitive through decreasing the cost of imported inputs and lowering the tax burden for exporters. It is also possible that sales tax on cars will be reduced or eliminated, which would result in an increase of domestic demand.	**Intentionality label:** IP intention. **Explanation:** Increasing access to foreign inputs.

Continued on the next page

#	Description (Continued)	Details (Continued)
15	On 6 July 2009, as a result of the 4th Green Growth Committee meeting, the government of the Republic of Korea announced several additional measures to back up the 'Green New Deal' program introduced earlier in the year. According to the press release, the government plans to raise fiscal support for R&D in the green industries to 2.8 trillion KRW (ca. 2.5 billion USD) by 2013 from 2.0 trillion KRW (ca. 1.8 billion USD) in 2009. Besides, the Korean Development Bank sets up a 300 billion KRW (ca. 273 million USD) fund for R&D and pre-market testing for these industries. The government also said that it will increase the established fund for SMEs in the green growth industries to 1.1 trillion KRW (ca. 1 billion USD) by 2013 from 60 billion KRW (ca. 54 million USD) in 2009. Furthermore, the credit guarantee scheme shall be more than doubled from 2.8 trillion KRW (ca. 2.5 billion USD) to 7 trillion KRW (ca. 637 billion USD) by 2013. In addition, it will provide an extended credit guarantee of 3 years for "green" startups with individual loan worth between 300 and 500 million KRW; 273,000-455,000 USD)...	**Intentionality label:** IP intention. **Explanation:** "Green IP policy" and R&D rules.
16	On 4 April 2016, the Russian government approved the Regulation on subsidies to aircraft engine manufacturers, covering coupon payments on the bonds issued by the latter under the guarantee of the Russian Government. The subsidies are issued to firms designing, manufacturing, testing and repairing aircraft engines, their components and associated instruments. The subsidies aim to cover coupon payments on the government-guaranteed bonds issued by the firms. The subsidies are supposed to cover the amount equal to 8% of annual interest on bonds over 20 million rubles and 6% on all other bonds. The document was amended with procedural modifications on 19 September 2017 (Decree N 1123), which did not alter the substance, extent or beneficiaries of the subsidy. The 2017 budget for the program was 2.4 billion rubles (USD 39.4m at the beginning of 2017), the 2018-2019 budgets - 2.7 billion for each year. The subsidies were distributed within Component 3 (Aircraft Engines) of the State Program for the Development of the Aircraft Manufacturing in 2013-2025.	**Intentionality label:** IP intention. **Explanation:** The name of the law is evidence of intent: "State Program for the Development of the Aircraft Manufacturing in 2013-2025".
17	The export loan was announced in November 2009 and finances the delivery of one Airbus A320-200 aircraft ('1 A 320-200') to the United States of America. The benefitting German exporter is Airbus S.A.S.. The German Eximbank only publishes value ranges for the projects it finances. The present project is in category 2. This category includes projects with a financing value between 16 and 50 million EUR. The GTA assumes the lower bound amount of the respective category, in this case 16 million EUR (24 million USD), as the conservative estimate of the project value. The maturity of the loan will be 10 years. The financing institution is BNP Paribas S.A., Paris.	**Intentionality label:** IP intention. **Explanation:** An export loan has, by our definition, industrial policy intentionality 'by construction'.

Table I.2: Other (Non-IP) Intentionality Examples

#	Description	Details
1	On 18 October 2019, the UK government added four medicines to the medicines that cannot be parallel exported from the UK list. The export ban was introduced on the subject goods due to anticipated shortages in the country. The list of medicines are: alprostadil, QVAR inhalers (beclometasone dipropionate), norethisterone and ranitidine. The medicines that cannot be parallel exported from the UK list was established in the beginning of October 2019 with twenty products subsequently being added, see related state act. The UK authorities noted that: "Parallel export of a medicine on the list (ed. medicines that cannot be parallel exported from the UK list) is considered a breach of regulation 43(2) of the Human Medicines Regulations 2012 and a contravention of the wholesale dealer license and may lead to regulatory action by the Medicines and Healthcare products Regulatory Agency (MHRA), which could include immediate suspension of the wholesale dealer licence."	**Intentionality label:** Other intention. **Explanation**: It's clear that the aim is to secure local availability of a certain good: "was introduced on the subject goods due to anticipated shortages in the country," which indicates a proximate cause other than an IP aim.
2	In May 8, 2019 Cambria Company LLC filed a petition seeking the imposition of AD and CVD orders against quartz surface products from India and Turkey. The products subject to the scope are currently classified in the Harmonized Tariff Schedule of the United States (HTSUS) under the following subheading: 6810.99.0010. Subject merchandise may also enter under subheadings 6810.11.0010, 6810.11.0070, 6810.19.1200, 6810.19.1400, 6810.19.5000, 6810.91.0000, 6810.99.0080, 6815.99.4070, 2506.10.0010, 2506.10.0050, 2506.20.0010, 2506.20.0080, and 7016.90.10. On June 24, 2019 the U.S. International Trade Commission (USITC) determined that there is a reasonable indication that a U.S. industry is materially injured or threatened with material injury by reason of imports of quartz surface products from India and Turkey that are allegedly subsidized and sold in the United States at less than fair value. On December 6, 2019 the U.S. Department of Commerce...	**Intentionality label:** Other intention. **Explanation:** This is an anti-dumping measure.
3	On 10 July 2020, the European Investment Bank (EIB) signed a EUR 100 million (approx. USD 113 million) multi-beneficiary intermediated loan (MBIL) agreement with Santander Consumer Finance SA to support SMEs and midcap companies. The credit line will more specifically support SMEs' and midcaps' investments in fleet acquisition and renewal including commercial fleet for land transport and agricultural machinery. According to the EIB: "The aim is to enhance access to finance of small/medium projects carried out by SMEs and midcaps." A state act in the GTA database is assessed solely in terms of the extent to which its implementation affects foreign commercial interests. On this metric, the investment support granted here is discriminatory.	**Intentionality label:** Other intention. **Explanation:** The objective is to support SMEs and midcap companies. This is a good example of employing selectivity. SME-s are being promoted without any selectivity in the types of SME-s the state is trying to foster.

Continued on the next page

#	Description (Continued)	Details (Continued)
4	The Japanese government announced on April 7th, 2020, an unprecedented emergency economic stimulus plan totalling USD 993bn (JPY 108.2tn) aimed at remedying the negative impact of the COVID-19 pandemic on the Japanese economy. One portion of this package was a fund of USD 2bn (JPY 220bn) budgeted for such firms wishing to move their operations back to Japan. The funding was provided to assist firms for whom supply chain issues arising from the COVID-19 outbreak were threatening their operations. Consistent media reports, as well as quotes from Japanese government officials, assert that the funding is primarily targeted towards moving such firms' operations out of China, due to the effects of the virus. The funding will be available to Japanese firms, and has no limitation depending on the industrial sector in which the firm operates.	**Intentionality label:** Other intention. **Explanation:** Response to COVID-19. Here, we have a short-term response to shocks, as opposed to longer-run goals of IP.
5	On 1 December 2018, the French authorities issued Decree No. 2018-1057 extending the scope of foreign investments in certain sectors subject to prior authorization. Such sectors are: "1) space operations; 2) cybersecurity; 3) artificial intelligence; 4) robotics; 5) semiconductors and additive manufacturing; 6) data hosting; 7) systems utilized for capturing computer data or intercepting correspondence; 8) IT systems for public authorities in the field of national security; 9) information systems utilized in crucial industries; 10) research and development of dual-use goods and technologies". In this context, the French authorities stipulated in said Decree that: "The decree extends the scope of the sectors covered by the authorization procedure to new economic sectors essential to guaranteeing the country's interests in matters of public order, public security or national defense." (own translation) The Decree entered into force on 1 January 2019.	**Intentionality label:** Other intention. **Explanation:** Addressing a non-industrial issue (public order, public security, national defense).
6	On 2 August 2013, the Clean Energy Finance Corporation (CEFC) of Australia announced the allocation of AUD 40 million (circa USD 36 million) to Sundrop Farms for the development of a solar thermal technology greenhouse complex near Port Augusta, South Australia. The 20-hectare greenhouse facility is expected to produce 15,000 tonnes of tomatoes per year. Clean Energy Finance Corporation The Clean Energy Finance Corporation is a government-funded financier created to fund clean energies. To achieve these goals, the Australian government has provided credits of AUD 2 billion each year from 1 July 2013. The statutory text, the Clean Energy Finance Act 2012, states that only solely or mainly Australian-based investments are eligible for these resources.	**Intentionality label:** Other intention. **Explanation:** Policy related to the development of sustainable farming.

Continued on the next page

#	Description (Continued)	Details (Continued)
7	On 18 November 2008 the Malaysian government adopted the document 'Measures To Address Impact Of Global Economic Slowdown On Malaysia's Trade And Industry', which provides for the full import duty exemption on raw materials and intermediate goods. The Ministry of International Trade and Industry considered these items 'productive imports' and the objective of duty exemptions is to reduce the cost of doing business for the manufacturing and construction sectors. Duties ranging from 5-30 percent on over 400 products were also eliminated. The products include iron and steel products such as steel bar and wire rods; petrochemicals and chemicals such as PVC, plastics, films and sheets, polyethylene and high impact polystyrene; textile and apparels such as man-made textile materials and textile fabrics covered with polyvinyl chloride; machinery and equipment such as moulding patterns made of plastic, wood or aluminium. International Trade Minister Muhyiddin Yassin stated the reason for introducing these measures is to help exporters avoid a slowdown and to boost construction industry.	**Intentionality label:** Other intention. **Explanation:** Name of the measure makes it clear that the measure is responding to short term business cycle fluctuation.

Table I.3: Not Enough Enough Information

#	Description	Details
1	On 6 October 2015, the Collegium of the Eurasian Economic Commission eliminated the export licensing requirement in relation to precious stones and metals, including those contained in various luxury goods and complex equipment (e.g. ballpoint pens, jewelry and optical instruments). Previously such exports were subject to licensing.	**Intentionality label:** Not enough information. **Explanation:** Entry does not stipulate any intention, it's only describing the measures.
2	In 2010, the Shangai-listed company SMEIC disclosed in its stock exchange filings the receipt of approximately 103 million USD of government subsidies. This represented a significant increase in the government subsidies obtained in the prior year (73 million USD). In China it is a legal requirement that publicly-listed firms report any subsidies received from government bodies. The stated subsidy amounts refer to a given calendar year. It is possible that this firm received other forms of state aid that have not been declared. The affected products and sectors have been chosen based on the "main products" and CSRC sector classification reported in the financial data of the firm. The subsidy amounts in USD were computed using the year-average exchange rate to the Chinese Yuan...	**Intentionality label:** Not enough information. **Explanation:** It's not clear what the intention of the government is.
3	The EIB will provide a 220 million EUR loan to support the construction and operation of a portfolio of eight wind farms and two solar photovoltaic (PV) power plants in France totalling 181MW by the Valeco Group.	**Intentionality label:** Not enough information. **Explanation:** Although this could be possible that this is another intention since it seems related to the environment; there is not enough information, however.

Continued on the next page

#	Description (Continued)	Details (Continued)
4	On June 13, 2017 the governor of Nevada signed into law a bill (AB280) that creates a preference of 5% for a bid or proposal for a state purchasing contract that is submitted by a Nevada-based business. To qualify for this preference, a business must certify that: (1) its principal place of business is in Nevada; or (2) a majority of the goods provided for in a state purchasing contract are produced in Nevada. The law prohibits granting the preference for the award of any contract which uses federal money, unless such a preference is authorized by federal law or any contract which has been procured on a multistate basis.	**Intentionality label:** Not enough information. **Explanation:** This is a good example of a local content requirement and preferential procurement where there is not sufficient evidence of intent. Neither is any explicit intent stated, nor is the measure selective in the types of producers or industries it wishes to promote.
5	Switzerland decreased the average tariff rate of 42 6-digit HS product categories within the GSP tariff regime in 2013 compared to the previous year available in the WTO Tariff Download Facility.	**Intentionality label:** Not enough information to determine. **Explanation:** There isn't an explicit intention of the measure.
6	Effective 07 Dezember 2009, the export reference priced was changed from 2290.54 RM/tonne to 2375.58 RM/tonne. The implied export tariff at the reference price changes from 20.4 percent to 20.74 percent.	**Intentionality label:** Not enough information. **Explanation:** No intent stated.
7	The credit for this transaction is provided by the Export Import Bank of India (EXIM) and requires that at least 75% of the contract price for goods and services associated with the project is sourced from India.	**Intentionality label:** Not enough information. **Explanation:** No intent stated. Even though the EXIM bank is providing the loan, it is not clear this is to finance an export transaction. Nor is there any selectivity applied in the local content requirement.
8	On 23 March 2016, the Indian Ministry of Commerce & Industry liberalized the FDI policy in the insurance sector by allowing automatic approval for consolidated foreign investments up to 49% of the equity capital of the insurance company. Earlier automatic approval was capped at 26% of the equity and investments above 26% and up to 49% required express approval from the government. The FDI limit has been kept unchanged for the insurance sector at 49%.	**Intentionality label:** Not enough information. **Explanation:** How the liberalization of the FDI policy ties in with broader policy objectives is not clear from the text.
9	At the same time, the amendment introduces the requirement that any foreign direct investment will previously require a confirmation by the Central Bank of Iceland.	**Intentionality label:** Not enough information. **Explanation:** No intention stated.

2. *Decision Flow Chart.* The below decision chart offers you a structured way to evaluate each entry. For a full version of the chart as a PDF click here.

Note that many of the nodes refer to the list of examples in the previous section 1.

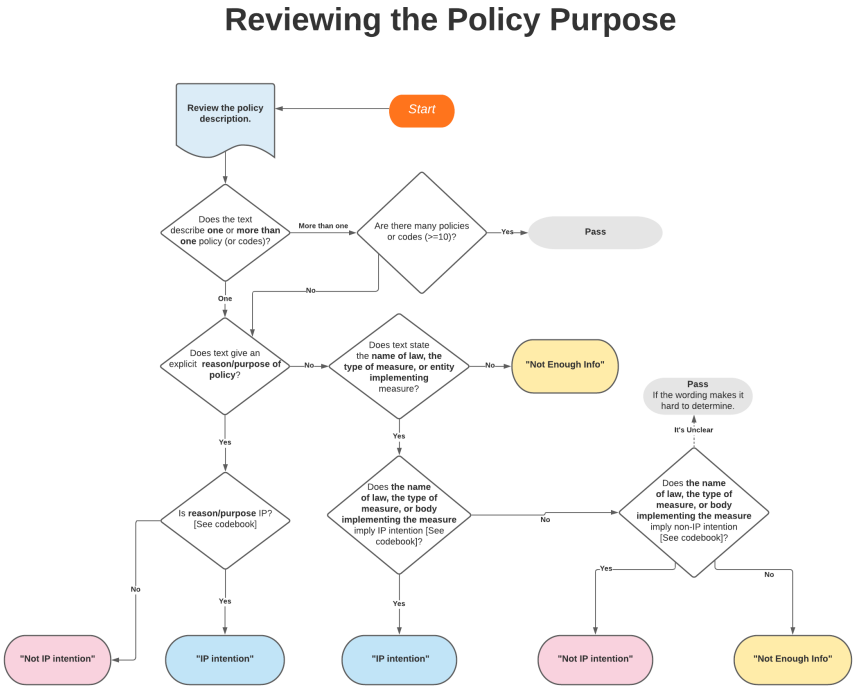


Figure I.2: Intentionality Coding Diagram