Patents as Indicators of Innovation: Large Language Models Meet Ground Truth in the LCD Industry

Mitsuru Igami* Shoki Kusaka† Simon Scheidegger‡ Anna Smirnova§

March 26, 2026

Abstract

Patents are widely used to measure innovation, but their economic content is ambiguous. We assess the usefulness of patent-based measures as indicators of the nature, incidence, and welfare impacts of innovation by leveraging a unique empirical context of the liquid crystal display (LCD) industry. We combine all US patents on LCD technology with detailed data on new products and manufacturing processes at the firm-product-fab level, as well as the structural estimates of their welfare impacts. Conventional statistics such as patent counts and citations fail to predict these ground truth events, whereas recently proposed text-based measures (e.g., novelty) exhibit some explanatory power. To extract further information from the text data, we use large language models (LLMs) to characterize and classify each patent based on its technological content and economic purpose. The results suggest LLM-based measures more than double the usefulness of patents as economic indicators.

1 Introduction

Patent statistics are ubiquitous in economics, but whether they capture the true economic content of innovation remains unsettled. The pioneers of this literature have long cautioned against their misuse, warning that their statistical and institutional linkages can be biased in many empirical contexts (Griliches, 1990; Lerner and Seru, 2022). Relying solely on raw

*University of Toronto
†Yale University
‡University of Lausanne
§University of Lausanne

counts or forward citations often conflates strategic portfolio growth with genuine inventive output, leaving the exact economic mechanisms and physical timing of these innovations ambiguous.

Our contribution to resolving this ambiguity is twofold. First, we leverage a concrete empirical context in which product innovations, process innovations, and the resulting welfare gains are reliably measured using physical ground truth and structural estimates. Second, we propose the expert-supervised deployment of large language models (LLMs) to carefully characterize the textual content of each patent document, instructing the models with precise economic definitions to extract specific technological mechanisms.

We revisit this question in the global LCD industry from 2001:Q1 to 2011:Q4. This setting features rapid product turnover and capital vintage upgrades (fab generations), where product and process innovations are directly observed. Furthermore, previous research demonstrates substantial welfare gains resulting from these innovations. We use three primary sets of ground-truth data: firm-product-quarter prices, quantities, and characteristics; fab order, installation, and ramp dates, alongside generations and yield/cost trajectories; and an estimated demand-supply system delivering welfare gains from innovation (that is, changes in buyer surplus, producer surplus, and total surplus caused by each of these product and process innovations).

To evaluate the predictive power of patents against this ground truth, we assemble a corpus of all LCD-related patents filed with the USPTO. While we construct conventional features such as counts, forward citations, and existing text-based metrics, these traditional tools often struggle to capture the nuanced engineering realities of the inventions. Therefore, we utilize LLMs to bridge the gap between unstructured patent text and structured economic data, extracting specific features that reflect the actual physical improvements and intended manufacturing goals described by the inventors.

We evaluate the usefulness of patents as indicators of economic innovations by addressing three interconnected questions regarding the nature, incidence, and magnitude of innovation. First, we instruct the LLMs to map the nature of innovation by distinguishing between product and process innovations and extracting finer engineering mechanisms, such as larger panel size, higher resolution, backlight shifts, or yield engineering. Second, we test whether these novel measures predict the incidence and timing of ground-truth product launches and process upgrades better than conventional counts and citations. Third, we assess the magnitude of their economic impact by identifying which patent-based measures are most closely associated with the ultimate welfare gains.

We find that conventional statistics such as patent counts and citations fail to predict ground truth events robustly, often requiring fully distributed lag models to account for severe persistence. While recently proposed text-based measures exhibit some explanatory power, our LLM-based classifications significantly outperform them. Specifically, patents classified by the LLM into specific product quality and manufacturing efficiency goals more than double the explanatory power for both the incidence of physical innovations and their contemporaneous welfare effects.

These methodological improvements have direct implications for the broader field of economics. Patents serve as the primary measure of innovation for all practical purposes in both economic research and public policy, including industrial, innovation, and competition policy. By refining the extraction of information from patent texts using expert-supervised LLMs, we can rejuvenate empirical innovation research. This approach directly improves the fundamental metrics that serve as the supporting evidence for formulating these policies and conducting their ex-post evaluations.

The remainder of the paper proceeds as follows. Section 2 clarifies our contribution within the existing literature on patent statistics, text-based measurement, and the use of LLMs in economics. Section 3 introduces the ground-truth data on LCD product launches, fab investments, and welfare estimates. Section 4 details the construction of the patent features, comparing traditional counts and recent text metrics before introducing our expert-supervised LLM classification schema. Section 5 evaluates the predictive performance of these measures against the physical incidence of product and process innovations. Section 6 assesses their ability to predict ultimate economic welfare gains, and Section 7 concludes.

2 Related Literature

Patent Statistics as Innovation Indicators. Patent counts and citations correlate with R&D activities and stock prices, and have been widely used as indicators of innovation (Hall et al., 2005; Bloom et al., 2013; Kogan et al., 2017). The pioneers of this literature have repeatedly cautioned against their misuse, warning that their statistical and institutional linkages can be biased in many empirical contexts (Hall et al., 1986; Griliches, 1990; Hall and Ziedonis, 2001; Shapiro, 2001; Ziedonis, 2004; Lerner and Seru, 2022). Our overall contribution is to use a concrete data context with the ground truth to propose the expert-supervised deployment of LLMs as the way forward in improving the usefulness of patent data.

Text-Based Innovation Measurement. A more recent literature replaces count-based statistics with text-based measures. Their work reflects a broader trend in economics to use text as data (Gentzkow et al., 2019). For instance, Kelly et al. (2021) construct scores based on the term frequency-inverse document frequency (TF-IDF) of patent texts, showing these metrics correlate with citation rates and firm market valuations at long horizons. We use Kelly et al.'s novelty, impact, and breakthrough indicators directly as one of the groups of predictors in our horse race regressions. Our central finding is that even though these text-based measures improve on raw counts and citations, our LLM-based measures substantially outperform them.

Large Language Models as Economic Measurement Tools. Large language models are rapidly becoming a popular tool kit for economic measurement (Korinek, 2023). Ludwig et al. (2025) distinguish between using LLMs for prediction versus measurement, noting that valid empirical analysis requires combining LLM outputs with external validation samples. Our empirical design implements such a validation logic by evaluating LLM outputs against observed physical events. In a related corporate finance application, Jha et al. (2024) prompt LLMs to extract anticipated investment changes from earnings calls, successfully predicting continuous firm-level actions. Similarly, Lopez-Lira and Tang (2023) and Chen et al. (2023) demonstrate that LLMs can perform structured classification of financial text, often outperforming simpler natural language processing approaches. We find that constraining model responses to an expert-curated label set helps extract genuine economic structure. Within the specific domain of patents, Yang (2023) applies ChatGPT to measure patent textual quality and predict application success as well as private patent value. We depart from this focus on private value and approval outcomes by instructing the LLMs to extract physical mechanisms that are grounded in economic theory (i.e., specific channels of product and process innovations) and linking these classifications to social welfare gains.

Patents Linked to Products and Processes. Finally, our work builds on efforts to map patents to specific real-world outcomes. de Rassenfosse (2018) highlights the fundamental "notice failure" challenge of linking commercial products to their underlying patents, examining virtual patent marking as a firm-level solution. Igami and Subrahmanyam (2019) conduct another ground-truth exercise for the hard disk drive industry, finding limited predictive power of conventional patent statistics for observed product introductions. Ganglmair et al. (2022) introduce the PATCCAT database to classify patent claims into product and process innovations. Argente et al. (2025) develop text-linking methods to match individual

patents to consumer product launches. Our empirical setting extends their framework to process innovations—such as fab generation upgrades and yield engineering—which require a different conceptual mapping.

3 Data

This section explains our data and the patent-based measures in the existing literature.

3.1 Sources and Ground-Truth Measures

Table 1 gives an overview of the data. We begin by gathering all patents on LCD inventions from the USPTO database. The full patent corpus contains 15,964 US LCD patents with filing dates spanning 2001 to 2016. However, all panel regressions restrict the sample to the 2001:Q1 through 2011:Q4 study window (14,571 patents) to match the availability of our ground-truth data. The remaining 1,393 patents filed after 2011 are classified by the LLM but excluded from the panel regressions.

Table 1: Data Overview

Data Category	**Unit of Observation**	**Count**	**Description**
Patents	Patent	15,964	US patents on LCD inventions, 2001–2011
Product innovations	Event	678	Product launches
Process innovations	Event	158	Fab investments
Welfare gains	Firm-quarter	224	Welfare effects of innovations
Consolidated panel 1	Firm-quarter	912	All 33 firms
Consolidated panel 2	Firm-quarter	472	All 14 firms with complete data
Consolidated panel 3	Firm-quarter	292	7 major firms only

To establish a ground truth (GT) about physical innovation, we utilize two proprietary databases purchased from DisplaySearch: a sales database and an engineering cost model. Covering the 2001:Q1 to 2011:Q4 window, these databases allow us to explicitly define physical events. A product innovation is recorded in the firm-quarter when the first sale of a new, distinct product occurs (defined by manufacturer, size, resolution, and backlight). A process innovation is defined by the discrete phases of production line investments in each firm's fabs, tracking equipment generation and capacity.

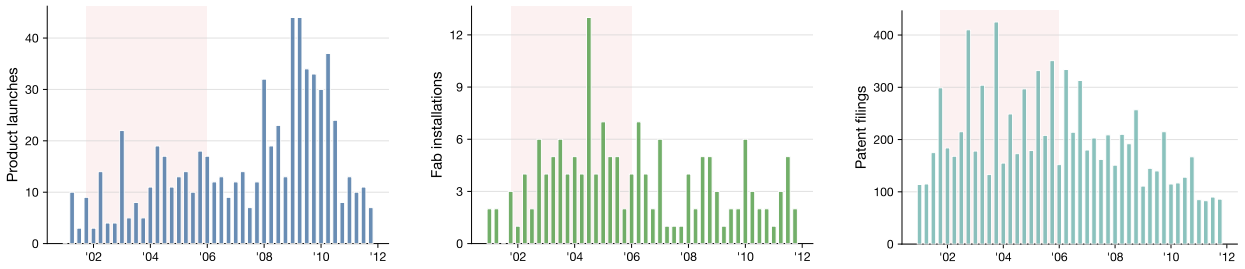
To quantify the ultimate economic impact of these events, we utilize structural welfare estimates from Igami et al. (2025). They estimate a differentiated product demand model for LCD panels across notebook, desktop, and television applications. By combining these

demand estimates with the DisplaySearch cost data, they conduct rigorous welfare counterfactuals to isolate the surplus contributions of specific events over their first four quarters. For product innovations, the counterfactual exercise eliminates the newly introduced product from the specific firm and quarter, simulates the alternative equilibrium outcomes for the subsequent four quarters, and calculates the difference in welfare from the baseline model. For process innovations, the counterfactual ignores the mass production ramp of the new fab, assuming the firm's cost structure has not improved from that specific upgrade. Re-equilibrating the market under these higher costs for the first four quarters yields the welfare gains specifically attributable to the manufacturing upgrades.

Finally, we merge the patent classifications, physical events, and welfare estimates into a consolidated panel dataset at the firm-quarter level. The overall dataset contains 33 firms that exist in both the patent data and the GT data, yielding 912 firm-quarter observations. A slightly more restricted subsample contains 14 firms (472 firm-quarters) which are shown to have both product and process innovations in the GT data. The most conservative subsample contains only the seven major firms that were known to be active in product and process innovations throughout the sample: specifically AUO, CMO, CPT, HannStar, LG, Samsung, and Sharp. This yields 292 firm-quarter observations. All but Sharp in these seven major firms participated in an international price-fixing cartel between 2001:Q4 and 2006:Q1, and this is the subsample that is highlighted by the shaded areas in all three panels of Figure 1.

3.2 Aggregate Patterns and Institutional Context

Figure 1 summarizes ground truth events and patent applications for 7 major firms, 2001:Q1 through 2011:Q4.



(A) Product Launches (B) Fab Installations (C) Patent Applications

Figure 1: Ground-Truth Events and Patents (7 Major Firms)[1]

[1]The shaded area covers 2001:Q4 to 2006:Q1 when the price-fixing cartel operated.

Panel A shows that the overall levels of new product introductions were elevated from 2004 as large LCD TVs were commercialized in East Asia. From around 2007, LCD TVs became mainstream not only in East Asia but globally, including North America and Western Europe. The final spike around 2009 and 2010 corresponds to the major technological shift from CCFL-based backlights to LED-based backlights.

Panel B shows the installations of new fab equipment in terms of the number of new production lines between 2003 and 2005 during an intense industry-wide capacity race. After 2006, weaker firms slowed down investments in new generation fabs, and therefore in terms of the number of new production lines, the new fab investments would seem to slow down. However, even though the dollar value and physical production capacity are not reported in the current draft or Figure 1, the dollar value and the physical production capacity of subsequent investments in even newer generations of equipment dramatically increased as new machinery became larger to handle larger input sheet glass and were much more expensive per unit. In this respect, whether the fab investments decreased over time or increased over time depends on the measure we use.

Panel C shows that patent filings are persistent but exhibit strong seasonality, typically peaking in Q2 and Q4, and therefore it seems likely to be driven by the semi-annual fiscal half-year cycle of corporate intellectual property review or patent application strategy. The overall filing levels shifted slightly downward after 2008.

The shaded area corresponds to the cartel (2001:Q1–2006:Q1). At a more granular level, we may partition the sample into six regimes: (i) dot-com bust (2001:Q1–2001:Q3), (ii) cartel working (2001:Q4–2004:Q3), (iii) cartel waning (2004:Q4–2006:Q1), (iv) TV boom (2006:Q2–2008:Q3), (v) Great Recession (2008:Q4–2009:Q2), and (vi) post-recession (2009:Q3–2011:Q4). In the future version, we plan to conduct analysis by regime.

A quick comparison of the three panels gives mixed impressions about their correlations. On the one hand, Panels A and C would seem to suggest a vaguely negative correlation between product innovations and patents. On the other hand, Panels B and C would seem to suggest a vaguely positive correlation between process innovations and patents. However, what further complicates the picture is that the dollar amount of fab investments as well as the physical capacity of output went up in later periods, even though they are not reported as data or studied in this draft. In the future version, we are planning to study how the result changes if we use the dollar amount of investment. In Section 5, we use firm-level panel regressions to formally investigate their relationships.

4 Patent-Based Measures

4.1 Traditional Count-Based Statistics

To set the baseline for evaluating the predictive power of patents, we construct conventional statistics, including simple counts and forward citations. We construct raw patent counts, which exhibit strong semi-annual and annual cycles, with the AR(2) and AR(4) lags dominating. This quarter-to-quarter mean reversion necessitates full distributed lag models to avoid severe omitted-variable bias. We also measure forward citations.

1: Patent Counts. For firm i in quarter t, we define the patent count as

$$\text{PatentCount}_{it} = \sum_{j \in \mathcal{P}_{it}} 1,$$

where $\mathcal{P}_{it}$ is the set of US LCD patents filed by firm i in quarter t.

2: Forward Citations. We measure forward citations received within a 5-year window of the grant date:

$$\text{FwdCit5yr}_{it} = \sum_{j \in \mathcal{P}_{it}} c_j^{(5)},$$

where $c_j^{(5)}$ counts citations received by patent j within 5 years of grant. The time window avoids the well-known truncation bias in raw citation counts. We also construct backward citation counts $\text{BwdCit}_{it} = \sum_{j \in \mathcal{P}_{it}} b_j$, where b_j is the number of prior-art references in patent j.

4.2 More Recent Text-Based Measures

We incorporate four sets of text-based measures proposed in the literature. First, we use TF-IDF metrics of novelty, impact, and breakthrough as proposed in Kelly et al. (2021). Second, we use the PATCCAT database constructed by Ganglmair et al. (2022), which classifies each patent and each claim into product innovation, process innovation, or a mixture of both based on specific lexicons or specific words that are used in the patent text. Third, we measure shifts in the firm's patent portfolio using the semantic embeddings of the text. Fourth, we construct lexicon-based indicators based on keywords that are related to specific physical improvements such as PPI, CCFL, LED, and ODF.

3: TF-IDF Measures. Following Kelly et al. (2021), we construct three patent-level scores based on TF-IDF representations of patent text. For patent j with term-frequency vector $\mathbf{tf}_j$, the TF-IDF representation is

$$\text{tfidf}_{w,j} = \text{tf}_{w,j} \times \log\left(\frac{N}{\text{df}_w}\right),$$

where N is the total number of patents and df_w is the number of patents containing term w.

Novelty measures how textually distinctive a patent is relative to the prior art at the time of filing:

$$\text{Novelty}_j = 1 - \max_{k \in \mathcal{B}_j} \cos(\mathbf{v}_j, \mathbf{v}_k),$$

where $\mathcal{B}_j$ is the set of patents filed in the 5 years before patent j, and $\mathbf{v}_j$ is the TF-IDF vector. *Impact* measures the similarity of a patent to subsequent work:

$$\text{Impact}_j = \max_{k \in \mathcal{F}_j} \cos(\mathbf{v}_j, \mathbf{v}_k),$$

where $\mathcal{F}_j$ is the set of patents filed in the 5 years after patent j. A patent is classified as a *breakthrough* if it scores in the top decile of both novelty and impact:

$$\text{Breakthrough}_j = \mathbf{1}\left[\text{Novelty}_j \geq q_{90}^{\text{nov}} \ \wedge \ \text{Impact}_j \geq q_{90}^{\text{imp}}\right].$$

We aggregate to the firm-quarter level as means (novelty, impact) or counts (breakthrough).

4: Patent Claim Classifications. Following the PATCCAT database of Ganglmair et al. (2022), each patent's claims are classified as product-oriented or process-oriented based on lexical patterns. We construct four firm-quarter aggregates:

$$\begin{aligned}
\text{ProductClaims}_{it} &= \sum_{j \in \mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ contains product claims}], \\
\text{ProcessClaims}_{it} &= \sum_{j \in \mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ contains process claims}], \\
\text{MixedClaims}_{it} &= \sum_{j \in \mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ contains both product and process claims}].
\end{aligned}$$

Mixed-claim patents account for 23.7% of the corpus.

5: Semantic Embedding Measures. We encode each patent abstract using the Sentence-BERT model all-MiniLM-L6-v2 (Reimers and Gurevych, 2019), producing a 384-dimensional embedding $\mathbf{e}_j \in \mathbb{R}^{384}$ for each patent j.

Portfolio shift measures how much a firm's patent portfolio changes direction between consecutive quarters:

$$\text{PortfolioShift}_{it} = 1 - \cos(\bar{\mathbf{e}}_{it},\ \bar{\mathbf{e}}_{i,t-1})\,,$$

where $\bar{\mathbf{e}}_{it} = \frac{1}{|\mathcal{P}_{it}|}\sum_{j\in\mathcal{P}_{it}} \mathbf{e}_j$ is the centroid of firm i's patent embeddings in quarter t. *Portfolio novelty* measures distance from the firm's own historical centroid, and *angular velocity* captures the rate of directional change.

6: Keyword-Based Mechanism Counts. We construct counts of patents mentioning specific LCD mechanisms using regular-expression matching against patent text:

$$\begin{aligned}
\text{PPI}_{it} &= \sum_{j\in\mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ mentions pixel-per-inch improvements}],\\
\text{LED}_{it} &= \sum_{j\in\mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ mentions LED backlight technology}],\\
\text{CCFL}_{it} &= \sum_{j\in\mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ mentions CCFL backlight technology}],\\
\text{ODF}_{it} &= \sum_{j\in\mathcal{P}_{it}} \mathbf{1}[\text{patent } j \text{ mentions one-drop-fill process}].
\end{aligned}$$

These capture specific physical improvements that an LCD industry expert would associate with product innovation (PPI, LED vs. CCFL transition) or process innovation (ODF).

4.3 Correlations among Conventional Measures

To bridge the gap between the mathematical definitions of these existing metrics and their actual predictive performance, we visualize their pair-wise Pearson correlations in Figure 2. Examining these underlying statistical patterns reveals several structural limitations in conventional patent measurement and foreshadows their behavior in our subsequent regressions.

First, the most prominent feature of the correlation matrix is a dense block of high positive correlations among the foundational patent statistics. Raw patent counts are heavily correlated with forward citations ($r = 0.89$) and backward citations ($r = 0.64$). Rather than capturing distinct leaps in technological quality or economic value, these metrics largely move together. They appear to capture a single latent variable—most likely a firm's overall scale

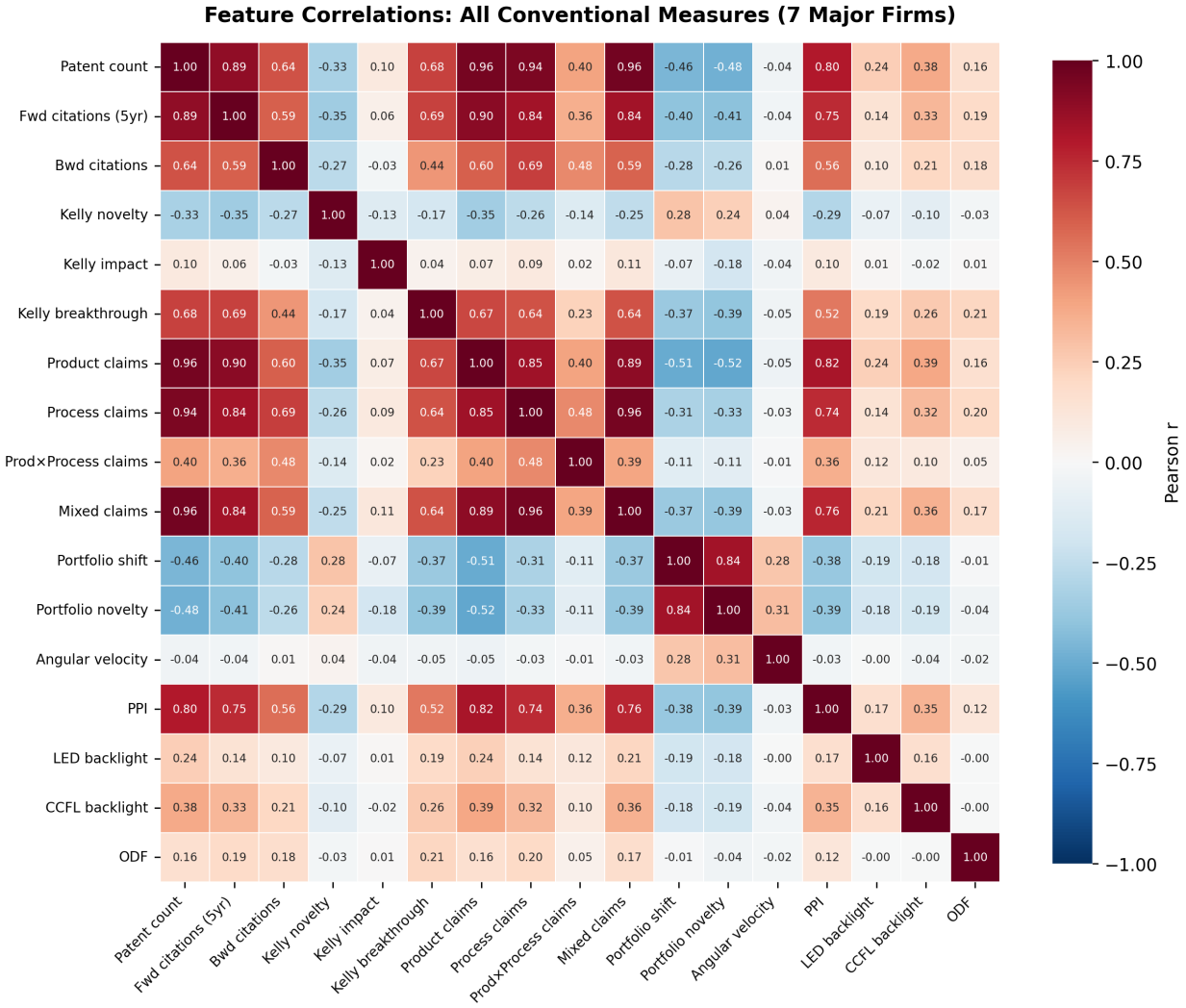


Figure 2: Pairwise correlations among Conventional Patent Measures

or general propensity to patent—meaning that high-volume "patent factories" mechanically generate high citation counts simply by virtue of their size.

Second, this "volume confound" severely impacts dictionary-based text approaches as well. The PATCCAT measures for product claims, process claims, and mixed claims are almost perfectly correlated with raw patent counts ($r = 0.94$ to $r = 0.96$). Furthermore, the correlation between product claims and process claims themselves is exceptionally high ($r = 0.85$). Because LCD patents routinely mention both the physical apparatus and the method of manufacturing it within the same document, simple keyword matching fails to separate the primary economic goal of the invention. Ultimately, these lexicon counts act as yet another proxy for firm scale rather than isolating specific types of innovation.

Third, in contrast, the Kelly et al. text metrics demonstrate the value of measuring semantic deviation rather than raw frequency. The Kelly novelty measure is negatively correlated with the dominant "volume" block, showing inverse relationships with both patent counts ($r = -0.33$) and forward citations ($r = -0.35$). By mathematically penalizing text that is overly similar to the existing corpus, this novelty metric successfully decouples itself from the patent factory scale, allowing it to capture genuinely distinct off-trend variations in a firm's research output.

Fourth, the vector-space embedding measures follow a similar structural pattern to the Kelly metrics. Portfolio shift and portfolio novelty are highly correlated with each other ($r = 0.84$) but are negatively correlated with raw patent volume ($r = -0.46$ and $r = -0.48$, respectively). By measuring the angular distance between a firm's current patent portfolio and its past portfolio or the overall industry, these metrics successfully filter out raw output levels to focus purely on the directional change of a firm's technological trajectory. Angular velocity, measuring the rate of this shift, is almost entirely orthogonal to every other measure in the matrix.

Finally, simple regular-expression counts of specific engineering terms prove to be either too generic or too noisy. The count of "PPI" (pixels per inch) is highly correlated with overall patent counts ($r = 0.80$), indicating that general industry terms simply scale mechanically with firm size. Conversely, highly specific technological keywords like "LED backlight," "CCFL backlight," and "ODF" exhibit near-zero correlations with almost every other metric. This extreme sparsity suggests that rigid keyword matching is often too fragile to consistently track aggregate firm-level economic outcomes across a decade of technological evolution.

4.4 Extracting Technological and Economic Content with LLMs

We extract further information from the patent text by using LLMs. As part of the prompt to the LLMs, we provide the following three contents: (1) the economic definitions of innovations, (2) the framework for classification of each patent, and (3) examples of expert classifications.

First, we clarify the economic definitions of product and process innovations. Product innovations improve the specifications of the final good and the buyer's willingness to pay. Key drivers include visual quality, response time, form factor, power, and reliability. Process innovations reduce the effective cost of manufacturing a given unit of quality. Key drivers include yield, throughput, materials, steps, and repair. As a tiebreaker between product and process innovations, if a structure exists solely to reduce manufacturing steps, it is a process innovation rather than a product innovation. Mixed innovations simultaneously achieve a significant product specification improvement and a significant manufacturing cost reduction. Because the LLMs' baseline training does not always adhere to strict economic concepts, it is important to clarify these definitions based on economic theory so that the LLMs will override any other definitions (e.g., legal or engineering ones).

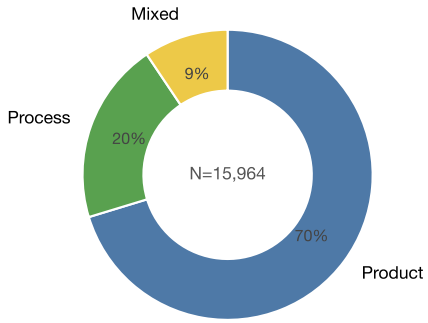
Second, we define the framework for the classification of each patent. We specifically prepared five slots to be filled. Slot 1 is the type, which is a choice between product, process, or mixed innovations as defined above. Slot 2 is the specific physical component that the patented invention improves, including thin-film transistor (TFT) array, liquid crystal (LC) panel, backlight unit, driver circuitry, and module assembly. Slot 3 is the product quality goal for the product innovation, to be chosen from the 11 categories: brightness, viewing angle, response time, power efficiency, form factor, reliability, visual uniformity, contrast ratio, color accuracy, resolution, and image quality. It must be N/A if the patent is process innovation. Slot 4 is the manufacturing efficiency goal, to be chosen from the five categories: yield, steps, materials, throughput, and repair. It must be N/A if the patent is product innovation. Mixed innovations will have both slots 3 and 4 filled. Slot 5 is the technological details, which is a natural language summary of the specific technology that has no predefined categories.

Third, the final component of the prompt instructions is the set of 46 patents chosen from the 100 patents that were classified by a human expert (Igami) in close collaboration with Gemini 3 Pro. This joint classification process was conducted iteratively to make sure the instructions were working and that the categories successfully captured the diverse patents present in the overall dataset. These 100 patents, jointly classified by Gemini 3 Pro and

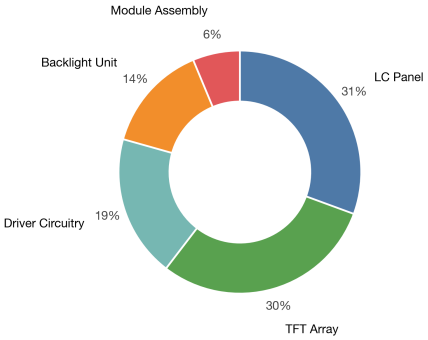
the final set of instructions under the supervision of the human expert, serve as the ground truth for classifying all of the LCD-related US patents in our dataset.

Because we cannot use human expert curation or supervision for the classification of every patent, we subsequently automated the prompting process. We tested whether the automated prompting would match the genuine joint classifications of the first 100 patents. The success rate was 92.0% for type classification (Product/Process/Mixed) and 73% for exact match across all five slots, evaluated against 200 expert-labeled patents. We further tested whether alternative LLMs would be able to match the same classification results. Claude Opus 4.6 achieved 89.5% type accuracy and 68.5% exact match; Gemini 3 Flash achieved 78.5% type accuracy and 63% exact match. An ensemble of Gemini 3 Pro and Claude Opus 4.6 (using the classification on which both models agreed, covering 90% of patents) achieved 96.1% type accuracy. Gemini 3 Pro in automated mode thus closely replicated the jointly classified patents under human supervision, whereas other LLMs performed slightly worse. The details of the classification schemes, as well as these 46 patents and the complete prompt, are available in the appendix and the replication package.

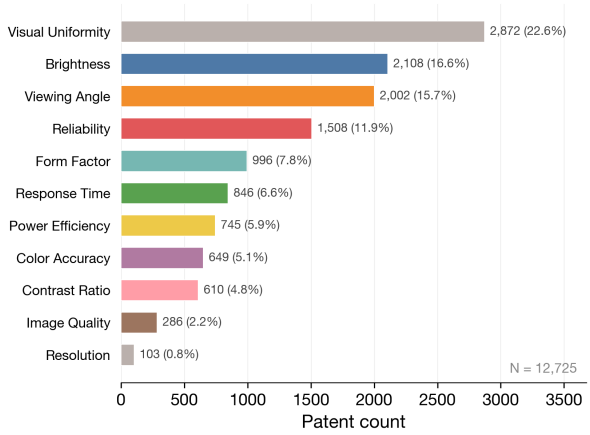
Figure 3 shows the automated classification results of all LCD patents by Gemini 3 Pro. Panel A shows the fractions of patents that are classified into product, process, and mixed innovations. Product innovations represent the vast majority of patents, whereas process and mixed innovations account for visible fractions as well (which remain mostly stable over time, though not reported here). Panel B shows the distribution of the five physical components on which the inventions were made (Slot 2). The LC panel and TFT array are the top two locations of patented inventions, followed by driver circuitry, the backlight unit, and module assembly. This makes sense because these first two are the most technologically dense components. Panel C reports the breakdown of product quality goals (Slot 3), which are more diverse than the other slots. Visual uniformity is the top category, followed by brightness and viewing angle. These are the three primary dimensions of image quality. Finally, Panel D shows that yield, materials, and steps are the three major categories of manufacturing efficiency goals (Slot 4). Yield improves the fraction of highly marketable, complete products. Materials refers to the reduction of material use, and steps refers to the reduction of manufacturing process steps. Along with the fourth category, throughput, these directly reduce the cost of waste or inputs. Repair is the category that captures methods to rescue defective outputs in the middle of the manufacturing process, indirectly helping to save inputs and costs as well.



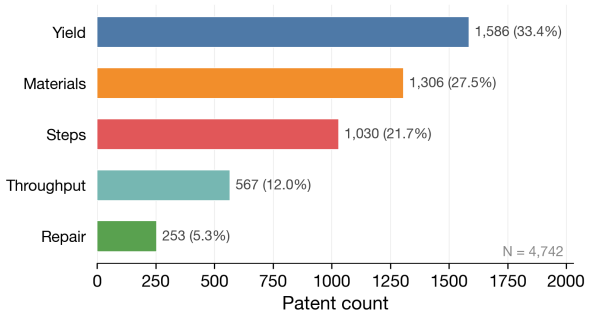
(A) Type of Innovation



(B) Component



(C) Product Quality Goal



(D) Manufacturing Efficiency Goal

Figure 3: LLM-Based Classification of 15,964 LCD Patents

4.5 Statistical Properties of the LLM-Based Measures.

As a preliminary check of the LLM-based classification, we conduct a falsification test predicting the ground truth (GT) outcomes. Table 3 summarizes the results of this preliminary test. To properly evaluate predictive performance, we compare regressions that share the same left-hand side variable. First, regarding product GT: product innovation patents successfully predict product GT, whereas process innovation patents should not predict product GT. Second, regarding process GT: process innovation patents successfully predict process GT, whereas product innovation patents should not predict process GT.

To further investigate how the LLM-extracted features interact with both the conventional baseline and with each other, we present a second correlation matrix in Figure 4. This visualization provides a transparent look at the statistical properties of the LLM classifications and foreshadows their performance in the subsequent incremental fit regressions.

Table 2: Falsification Test: LLM Patent Types vs. Ground Truth Outcomes

LHS variable	**RHS variable**	**Expected result**	**OLS R^2**	**F-test p value**
Product GT	Product patent count	Should predict	0.0789	0.027
Product GT	Process patent count	Should not predict	0.0335	0.552
Process GT	Process patent count	Should predict	0.0281	0.686
Process GT	Product patent count	Should not predict	0.0255	0.747

Note: The number of observations is 236 for all regressions. OLS with firm and time fixed effects, Cluster-Robust Variance 1 (CRV1) clustered standard errors. Lead/lag window is 4 quarters each. The Joint F-test tests the null hypothesis that all lead, contemporaneous, and lag coefficients are simultaneously zero. Product GT regressed on Product Patents, and Process GT regressed on Process Patents should predict (positive controls). Product GT regressed on Process Patents, and Process GT regressed on Product Patents should not (negative controls).

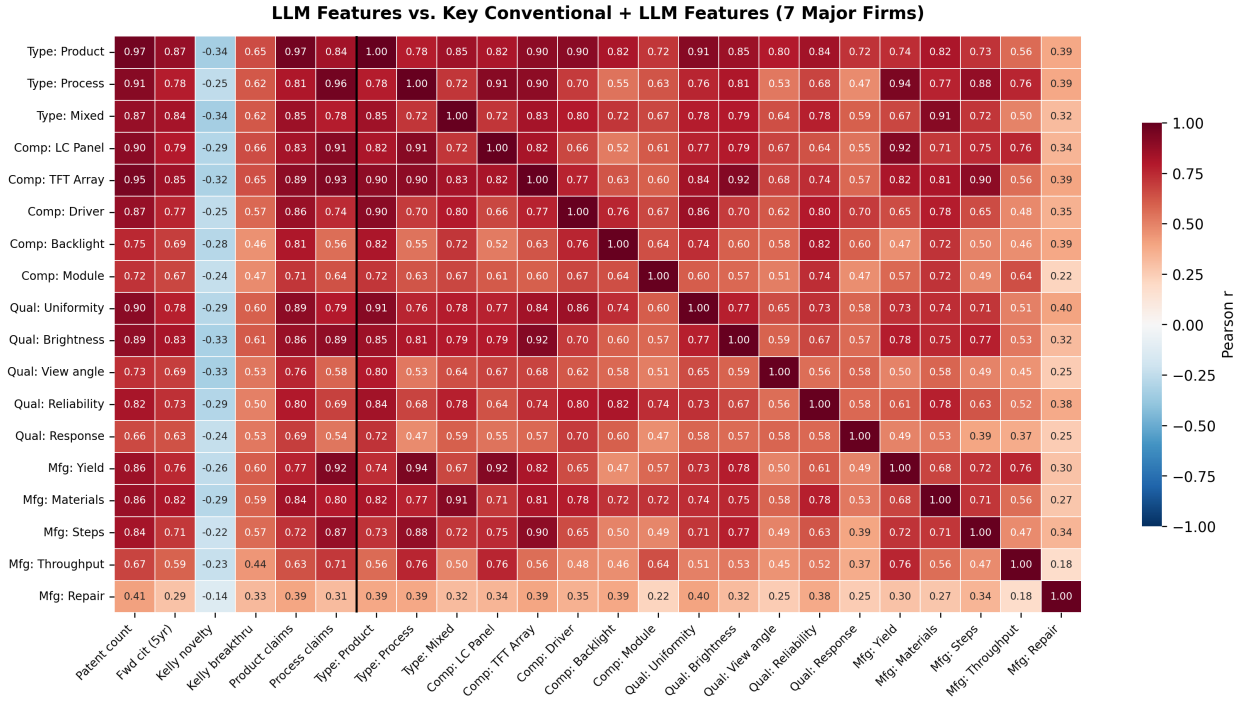


Figure 4: Pairwise Correlations of LLM Measures

First, looking at the intersection of the LLM features and the conventional metrics, it is clear that extracting specific engineering goals does not entirely eliminate the underlying firm-scale confound. Almost all major LLM features, such as the product type identifier and core components like the TFT array, remain heavily correlated with raw patent counts ($r = 0.85$ to $r = 0.97$). When aggregating at the firm-quarter level, large firms simply patent more of everything. Using an LLM to merely count "product patents" yields a metric that behaves similarly to a traditional volume measure.

Second, in contrast to the volume metrics, the Kelly et al. novelty measure shows per-

sistent negative correlations (ranging from $r = -0.24$ to $r = -0.34$) against almost every single LLM feature. This confirms that the LLM is capturing a fundamentally different signal from existing advanced text metrics. The Kelly metric captures how much a document's text deviates from the historical corpus, whereas the LLM strictly categorizes what the engineers are physically trying to achieve. Because these signals are mathematically orthogonal, they provide complementary information, explaining the substantial incremental jumps in adjusted R^2 when LLM features are added to baseline text models.

Third, examining the correlations among the LLM features themselves (the right side of the matrix) reveals dense blocks of co-movement. For instance, innovations in the LC panel and the TFT array are highly correlated ($r = 0.82$), and manufacturing yield correlates strongly with manufacturing materials ($r = 0.68$). This accurately reflects the physical reality of the LCD industry. When a firm invests in a massive new generation of fab, it executes a bundled, multi-front R&D campaign across multiple components and processes simultaneously. The LLM successfully recovers these bundled R&D strategies.

Fourth, the true value-add of the LLM extraction emerges in the sparse, niche engineering goals that break away from the dominant volume trend. Features such as manufacturing repair, throughput, and response time exhibit noticeably lower correlations with overall patent counts ($r = 0.41$ for repair) and weak correlations with other LLM features. Rather than lumping all process improvements into a single dictionary-based category, the LLM flags highly specific, episodic technical thrusts. These granular, independent vectors capture short-term engineering campaigns that do not simply scale with the overall size of the firm's patent portfolio, providing the distinct signals that ultimately map to concrete welfare gains.

5 Predicting Product and Process Innovations

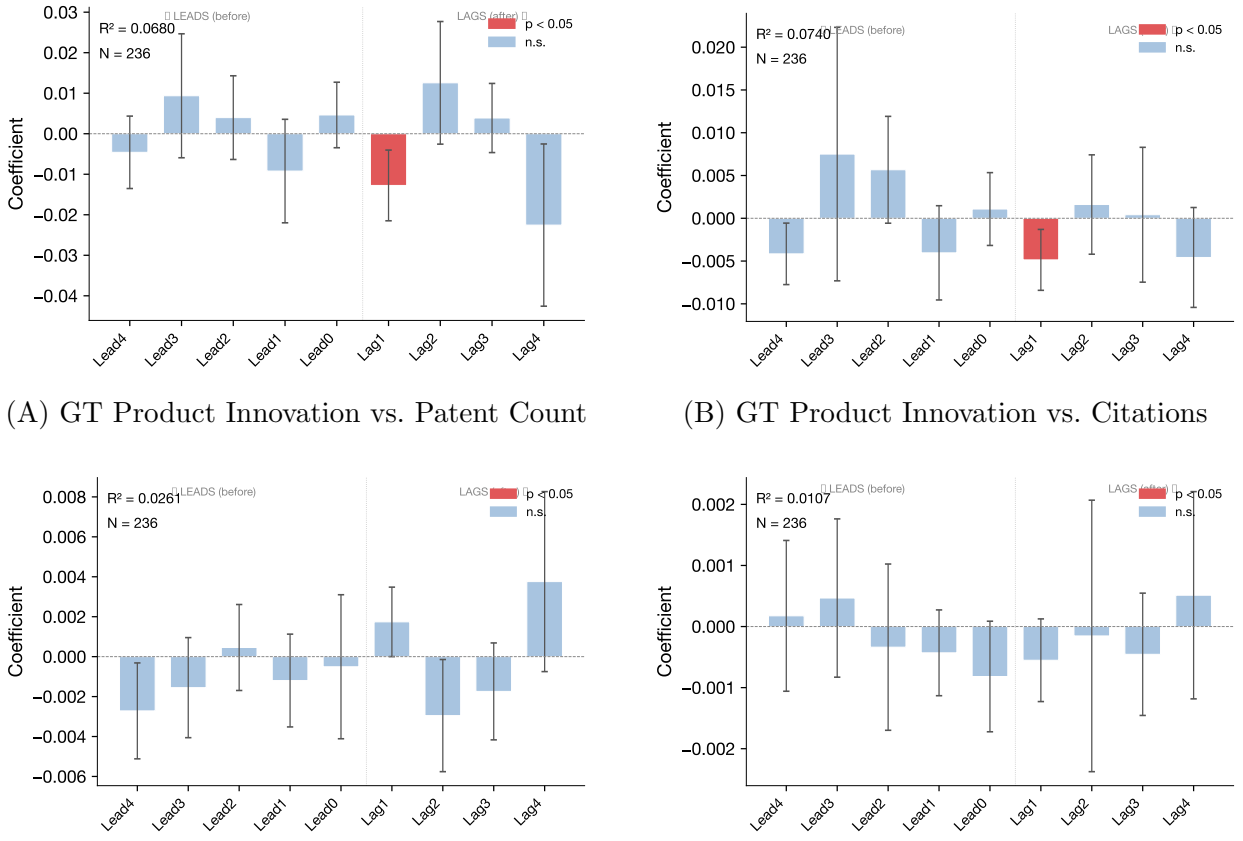
This section uses our consolidated firm-level panel data to systematically evaluate the usefulness of patent-based variables in predicting the incidence of product and process innovations in our ground truth (GT) data. Section 5.1 uses patent counts and forward citations with a variety of lead/lag structures. Section 5.2 introduces all other measures and compares their predictive performance both in isolation and in combination.

5.1 Patent Counts, Forward Citations, and Lead/Lag Structures

We start our analysis with the regression of the counts of product and process innovations in our GT data on the most basic patent statistics: counts and forward citations. Because

these left-hand side (LHS) variables are count data, we use Poisson and negative binomial models in addition to OLS. Whether patents are leading or lagging indicators of innovation is not obvious. Hence, our baseline specification includes not only the contemporaneous patent statistic (either count or citation), but also four quarters of its lags and four quarters of leads, $X_{i,t+l}$, where $l \in \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$. We also include fixed effects for both firms and time periods.

Figure 5 plots the coefficient estimates and their 95% confidence intervals for the basic patent statistics and their leads and lags. Panels A and B show that the only statistically significant variable to predict GT product innovations in period t is the patent count or citation in period $t + 1$, and its coefficient is negative. The combination of mostly insignificant coefficient estimates and the mixture of positive and negative signs suggests severe multicollinearity. Panels C and D reveal similarly inconclusive patterns when predicting GT process innovations.



(A) GT Product Innovation vs. Patent Count

(B) GT Product Innovation vs. Citations

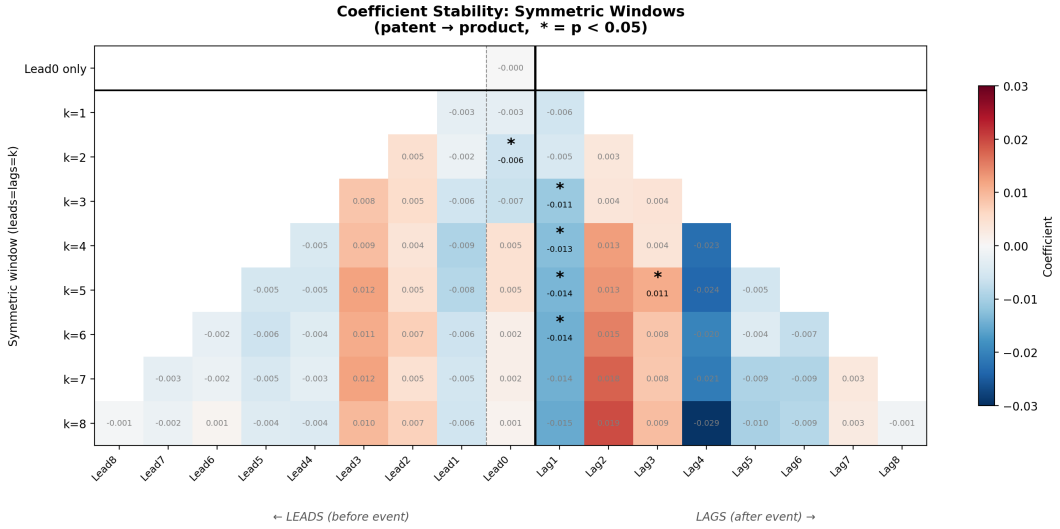
(C) GT Process Innovation vs. Patent Count

(D) GT Process Innovation vs. Citations

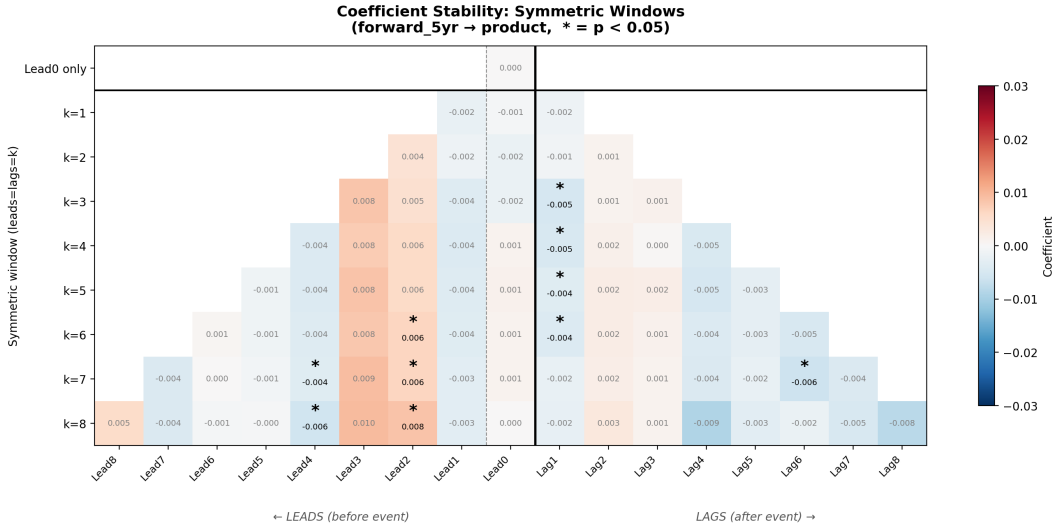
Figure 5: Regression of GT Events on Basic Patent Statistics with Leads and Lags

To expand and generalize this analysis, we systematically investigate other possible spec-

ifications of the lead/lag structure, ranging from using only the contemporaneous statistic, X_{it}, to the inclusion of up to eight quarters of leads and lags, $l \in \{-8, \ldots, 0, \ldots, 8\}$. Due to space constraints, we report these full coefficient stability plots in Appendix B. Figure 6 houses the two subfigures for product innovations regressed on patent counts and forward citations. Figure 7 presents the corresponding two subfigures for process innovations. Across all specifications, the only robust conclusion is that the correlation patterns are noisy and inconclusive. This finding resonates with that of Hall et al. (1986), who conclude that given the high serial correlation of patent statistics, the identification of their specific lead/lag structure is practically impossible.

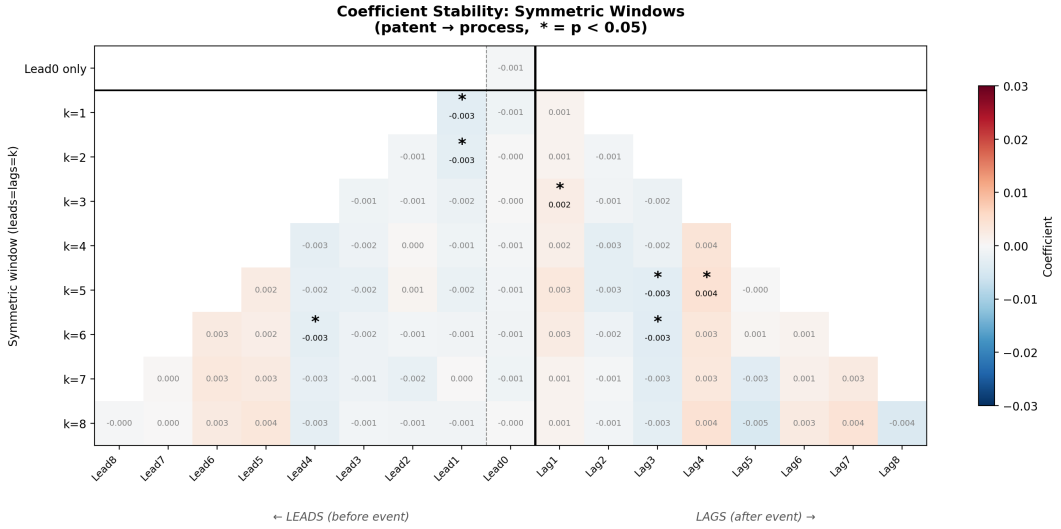


(A) Patent Count

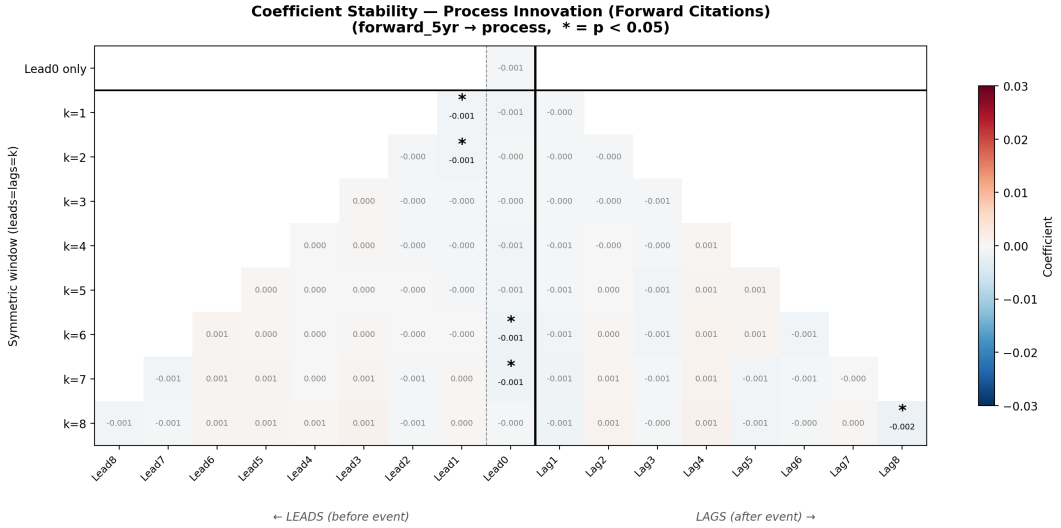


(B) Forward Citations

Figure 6: Sensitivity to Lead/Lag Length (Product Innovation)



(A) Patent Count



(B) Forward Citations

Figure 7: Sensitivity to Lead/Lag Length (Process Innovation)

5.2 Incorporating the Text-Based and LLM-Based Measures

We now regress the GT product and process innovations on not only the basic counts and forward citations, but also the more recently proposed text-based measures and our LLM-based variables. First, we study the predictive performance of each of the ten groups of regressors in isolation. Table 3 summarizes the in-sample fit of these independent regressions. Note that the numbers in parentheses next to the fit measures indicate the rank of each regressor group within that column, and boldface values highlight the top three rankers.

Table 3: In-Sample Fit (Product and Process Innovations)

LHS variable	Product innovation			Process innovation	
Regressor group	OLS R^2	Poisson $\tilde{R}^2$	NegBin $\tilde{R}^2$	OLS R^2	Poisson $\tilde{R}^2$
1. Patent count	0.0000 (9)	0.3047 (10)	0.1955 (9)	0.0006 (10)	0.1604 (10)
2. Forward citation	0.0000 (10)	0.3049 (8)	0.1956 (8)	0.0019 (9)	0.1609 (9)
3. Kelly text	**0.0132 (3)**	0.3072 (4)	0.1974 (5)	**0.0334 (1)**	**0.1765 (1)**
4. PATCCAT	0.0083 (5)	0.3070 (5)	0.1982 (4)	0.0078 (4)	**0.1632 (3)**
5. Semantic embedding	0.0063 (6)	0.3048 (9)	0.1955 (10)	0.0069 (5)	0.1620 (7)
6. Keyword count	0.0058 (7)	0.3054 (6)	0.1962 (6)	**0.0117 (2)**	**0.1642 (2)**
7. LLM type	0.0013 (8)	0.3053 (7)	0.1960 (7)	0.0021 (8)	0.1611 (8)
8. LLM component	**0.0159 (2)**	**0.3082 (3)**	**0.1991 (2)**	0.0067 (6)	0.1624 (5)
9. LLM product goal	**0.0456 (1)**	**0.3163 (1)**	**0.2071 (1)**	**0.0080 (3)**	0.1626 (4)
10. LLM process goal	0.0115 (4)	**0.3085 (2)**	**0.1990 (3)**	0.0064 (7)	0.1624 (6)

Note: Ranks are in parentheses, with top three values in bold. Each row reports the goodness-of-fit from an independent regression of the outcome on the indicated regressor group plus firm and time fixed effects. The number of observations is 292 in all regressions. OLS R^2 reports the partial R^2 attributable to the regressor group (beyond fixed effects). Poisson $\tilde{R}^2$ and NegBin $\tilde{R}^2$ report McFadden pseudo-R^2 for Poisson and negative binomial regressions, respectively. The latter for process innovation is not reported because the estimation algorithm fails to converge.

The basic measures of patent counts and forward citations hardly move the needle, ranking lowest among all ten groups of regressors for both product and process innovations. In contrast, recently suggested text-based measures in the literature achieve measurable contributions. For example, in predicting product innovations, the Kelly et al. text metrics achieve an OLS R^2 of 0.0132 (rank 3). For process innovations, these text-based measures are actually among the top performers. We do not compare the absolute magnitude of the fit measures between product and process innovations because their left-hand side variables are fundamentally different.

Finally, the LLM-based measures are among the top rankers regardless of the specific fit metric used for product innovations. The LLM product goal ranks first with an R^2 of 0.0456, followed by the LLM component at rank two with an R^2 of 0.0159. For process innovations, the LLM product goal ranks among the top predictors, but the other LLM groups show more mixed performance. Given the relative scarcity of process innovation instances in the

Table 4: Incremental Explanatory Power (Product and Process Innovations)

LHS variable	Product innovation				Process innovation			
Measure of fit	R^2		Adj R^2		R^2		Adj R^2	
Regressor groups	Level	Change	Level	Change	Level	Change	Level	Change
Patent count (baseline)	0.0000	—	−0.0034	—	0.0006	—	−0.0028	—
+ Fwd citation	0.0000	0.0000	−0.0104	−0.0070	0.0024	0.0018	−0.0079	−0.0051
+ Kelly text	0.0135	0.0135	−0.0072	**+0.0032**	0.0354	0.0330	0.0151	**+0.0230**
+ PATCCAT	0.0271	0.0136	−0.0075	−0.0003	0.0442	0.0088	0.0101	−0.0050
+ Semantic	0.0353	0.0082	−0.0098	−0.0023	0.0459	0.0017	0.0012	−0.0089
+ Keyword count	0.0413	0.0060	−0.0182	−0.0084	0.0586	0.0127	0.0002	−0.0010
+ LLM type	0.0459	0.0046	−0.0245	−0.0063	0.0588	0.0002	−0.0107	−0.0109
+ LLM component	0.0800	0.0341	−0.0064	**+0.0181**	0.0711	0.0123	−0.0162	−0.0055
+ LLM product goal	0.1204	0.0404	0.0193	**+0.0257**	0.0922	0.0211	−0.0121	**+0.0041**
+ LLM process goal	0.1272	0.0068	0.0079	−0.0114	0.1068	0.0146	−0.0153	−0.0032

Note: All results are based on OLS regressions, with firm and time fixed effects included. Each row adds the indicated regressor group cumulatively. The number of observations is 292 in all regressions. The changes report ΔR^2 and ΔAdj R^2 from the preceding row. Positive values of ΔAdj R^2 are in bold.

GT data, this result for process outcomes is not totally surprising.

6 Predicting Welfare Gains

To answer whether patents capture ultimate economic value, we substitute continuous welfare changes as the left-hand side outcomes, estimated strictly via OLS. Table 5 presents the standalone explanatory power of each regressor group for these welfare impacts. As in the previous table, parentheses indicate column rankings and boldface denotes the top three performers. Similar to the incidence results, conventional patent counts and citations explain minimal variance, yielding R^2 values near zero across all welfare measures. Existing text-based measures perform slightly better but remain constrained in their explanatory power; the Kelly et al. metrics, for instance, reach an R^2 of 0.0151 for absolute changes in total surplus. However, the LLM-extracted features are exceptionally predictive of welfare. Specifically, the LLM component and LLM product quality groups yield the highest standalone OLS R^2 values. For absolute changes in total surplus, the LLM component ranks first ($R^2 = 0.0712$), followed closely by the LLM product goal in second place ($R^2 = 0.0676$). The specific engineering slots extracted by the LLMs capture the actual economic magnitude of the innovations in a way that generalized text metrics do not.

Table 6 applies the same cumulative step-wise approach to evaluate the incremental explanatory power for welfare gains. As with the incidence models, we focus on the Adj R^2 to account for the expanding parameter space. When predicting absolute changes in total sur-

Table 5: In-Sample Fit (Welfare Gains)

LHS variable	Dollar values			Percentage changes		
Regressor group	ΔBS	ΔPS	ΔTS	%ΔBS	%ΔPS	%ΔTS
1. Patent count	0.0004 (10)	0.0027 (10)	0.0012 (9)	0.0022 (10)	0.0173 (8)	0.0077 (9)
2. Forward citation	0.0005 (8)	0.0073 (6)	0.0025 (7)	0.0087 (6)	0.0355 (4)	0.0195 (5)
3. Kelly text	0.0157 (4)	0.0103 (5)	0.0151 (4)	0.0175 (5)	0.0076 (9)	0.0194 (6)
4. PATCCAT	0.0097 (5)	0.0137 (4)	0.0108 (5)	0.0063 (7)	0.0234 (5)	0.0080 (8)
5. Semantic embedding	0.0005 (9)	0.0034 (9)	0.0004 (10)	0.0046 (9)	0.0046 (10)	0.0033 (10)
6. Keyword count	0.0052 (6)	0.0051 (7)	0.0052 (6)	0.0233 (4)	0.0175 (7)	0.0269 (4)
7. LLM type	0.0009 (7)	0.0050 (8)	0.0014 (8)	0.0051 (8)	0.0215 (6)	0.0131 (7)
8. LLM component	**0.0509 (2)**	**0.0931 (1)**	**0.0712 (1)**	**0.0418 (1)**	**0.0801 (3)**	**0.0444 (3)**
9. LLM product goal	**0.0674 (1)**	**0.0556 (3)**	**0.0676 (2)**	**0.0399 (2)**	**0.0814 (2)**	**0.0632 (1)**
10. LLM process goal	**0.0354 (3)**	**0.0592 (2)**	**0.0497 (3)**	**0.0308 (3)**	**0.1252 (1)**	**0.0592 (2)**

Note: Ranks are in parentheses, with top three values in bold. Each row reports the OLS R^2 (partial, beyond fixed effects) from an independent regression of the welfare outcome on the indicated regressor group plus firm and time fixed effects. The number of observations is 292 in all regressions. ΔBS, ΔPS, and ΔTS denote the changes in buyer surplus, producer surplus, and total surplus, respectively.

Table 6: Incremental Explanatory Power (Welfare Gains from Innovation)

LHS variable	ΔTS				%ΔTS			
Measure of fit	R^2		Adj R^2		R^2		Adj R^2	
Regressor groups	Level	Change	Level	Change	Level	Change	Level	Change
Patent count (baseline)	0.0012	—	−0.0023	—	0.0077	—	0.0042	—
+ Fwd citation	0.0025	0.0013	−0.0079	−0.0056	0.0224	0.0147	0.0123	**+0.0081**
+ Kelly text	0.0250	0.0225	0.0045	**+0.0124**	0.0518	0.0294	0.0319	**+0.0196**
+ PATCCAT	0.0418	0.0168	0.0077	**+0.0032**	0.0584	0.0066	0.0249	−0.0070
+ Semantic	0.0431	0.0013	−0.0017	−0.0094	0.0618	0.0034	0.0179	−0.0070
+ Keyword count	0.0498	0.0067	−0.0092	−0.0075	0.0781	0.0163	0.0209	**+0.0030**
+ LLM type	0.0516	0.0018	−0.0184	−0.0092	0.0825	0.0044	0.0148	−0.0061
+ LLM component	0.1372	0.0856	0.0561	**+0.0745**	0.1251	0.0426	0.0428	**+0.0280**
+ LLM product goal	0.1926	0.0554	0.0998	**+0.0437**	0.1637	0.0386	0.0675	**+0.0247**
+ LLM process goal	0.2161	0.0235	0.1089	**+0.0091**	0.2270	0.0633	0.1214	**+0.0539**

Note: All results are based on OLS regressions, with firm and time fixed effects included. Each row adds the indicated regressor group cumulatively. The number of observations is 292 in all regressions. The changes report ΔR^2 and ΔAdj R^2 from the preceding row. Positive values of ΔAdj R^2 are in bold. TS denotes total surplus.

plus (ΔTS), stacking the basic counts, citations, and text measures provides only marginal improvements; the Kelly et al. text features offer the only notable bump (+0.0124). The narrative shifts dramatically once the LLM features enter the model. Incorporating the LLM component group triggers a massive increase in the $\Delta Adj R^2$ (+0.0745). Subsequently adding the LLM product quality group further increases the fit (+0.0438). The fully saturated model explains over 10 percent of the variance in absolute total surplus changes (Adj $R^2 = 0.1089$) and over 12 percent of the variance in percentage changes (Adj $R^2 = 0.1214$). Crucially, the vast majority of this explanatory power originates entirely from the introduc-

tion of the LLM-extracted features, proving that they successfully map to the true economic magnitude of the innovations.

7 Conclusion

The overarching goal of this paper is to reassess the usefulness of patents as indicators of real innovations and their ultimate economic impacts. By leveraging the concrete physical ground truth and structural welfare estimates of the LCD industry, we subject both traditional and frontier patent metrics to a rigorous empirical test. At first glance, the overall explanatory power across our models might appear modest; even our fully saturated models explain roughly 10 to 12 percent of the variance in absolute welfare gains. However, this is a reflection of scientific reality rather than an empirical failure. Patents are inherently noisy, strategically driven legal documents. As our baseline results confirm, simple patent counts and forward citations hardly move the needle in predicting physical product launches, fab investments, or shifts in consumer and producer surplus.

However, this sobering baseline also highlights the vital progress made by recent literature. The transition from count-based statistics to text-based metrics represents a definitive step in the right direction. Existing text-based measures, such as novelty scores and lexicon-based classifications, successfully extract meaningful semantic signals from patent documents and achieve measurable contributions in predicting ground-truth events. Our results validate this methodological shift, confirming that the text of a patent holds far more economic value than its mere existence or its position within a citation network.

Our primary contribution is to show that we can significantly accelerate this progress by bringing economic theory directly into the extraction process. By utilizing large language models under expert supervision, we impose strict economic definitions—distinguishing between product quality goals and manufacturing efficiency goals—onto unstructured text. This structured approach yields quantitatively important improvements, often doubling or tripling the explanatory power of the best existing text-based measures, particularly when predicting ultimate welfare gains.

Ultimately, our findings do not suggest that LLMs are a panacea that perfectly resolves the ambiguity of patent data. The unexplained variance remains substantial. Rather, we demonstrate that the economics community's ongoing effort to extract richer information from patent texts is bearing real fruit. When concrete ground-truth data and rigorous economic frameworks are combined with modern LLMs, we can systematically bridge the

gap between legal filings and genuine economic indicators. This approach rejuvenates the empirical value of patent data, offering a more reliable foundation for future innovation research and policy evaluation.

References

Argente, David, Salomé Baslandze, Douglas Hanley, and Sara Moreira, "Patents to Products: Product Innovation and Firm Dynamics," *Federal Reserve Bank of Atlanta Working Paper*, 2025, (2020-4a).

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

Chen, Yifei, Bryan T. Kelly, and Dacheng Xiu, "Expected Returns and Large Language Models," *SSRN Working Paper*, 2023, (4416687).

de Rassenfosse, Gaétan, "Notice Failure Revisited: Evidence on the Use of Virtual Patent Marking," Working Paper 24288, National Bureau of Economic Research February 2018.

Ganglmair, Bernhard, W. Keith Robinson, and Michael Seeligson, "The Rise of Process Claims: Evidence from a Century of U.S. Patents," *ZEW Discussion Paper*, 2022, (22-011).

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

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 1990, *28* (4), 1661–1707.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, 2005, *36* (1), 16–38.

_ **and Rosemarie Ham Ziedonis**, "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995," *RAND Journal of Economics*, 2001, *32* (1), 101–128.

_ **, Zvi Griliches, and Jerry A. Hausman**, "Patents and R and D: Is There a Lag?," *International Economic Review*, 1986, *27* (2), 265–283.

Igami, Mitsuru and Jai Subrahmanyam, "Patent Statistics as an Innovation Indicator? Evidence from the Hard Disk Drive Industry," *The Japanese Economic Review*, 2019, *70* (3), 308–330.

_ , Shoki Kusaka, Jeff Qiu, and Tuyetanh L. Tran, "Welfare Gains from Product and Process Innovations: The Case of LCD Panels, 2001—2011," 2025. Working Paper.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, "ChatGPT and Corporate Policies," *NBER Working Paper*, 2024, (32161).

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Korinek, Anton, "Generative AI for Economic Research: Use Cases and Implications for Economists," *Journal of Economic Literature*, 2023, *61* (4), 1281–1317.

Lerner, Josh and Amit Seru, "The Use and Misuse of Patent Data: Issues for Finance and Beyond," *The Review of Financial Studies*, 2022, *35* (6), 2667–2704.

Lopez-Lira, Alejandro and Yuehua Tang, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," *arXiv preprint*, 2023.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan, "Large Language Models: An Applied Econometric Framework," *NBER Working Paper*, 2025, (33344).

Reimers, Nils and Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in "Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)" 2019, pp. 3982–3992.

Shapiro, Carl, "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting," in "Innovation Policy and the Economy," Vol. 1, MIT Press, 2001, pp. 119–150.

Yang, Stephen, "Understanding Innovation Quality and Success with Large Language Models," *SSRN Working Paper*, 2023, (4482536).

Ziedonis, Rosemarie Ham, "Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms," *Management Science*, 2004, *50* (6), 804–820.

Appendix: LLM Classification Protocol

We provide the complete system prompt used to instruct Gemini 3 Pro for patent classification. The prompt defines a 5-slot schema grounded in economic theory, distinguishing product innovations (raising consumer willingness-to-pay) from process innovations (reducing manufacturing costs). A "Golden Set" of 46 expert-classified example patents (not shown here) accompanies the prompt as few-shot demonstrations.

Core Economic Logic

1. **Product Innovation ("Raising the Ceiling"):** Innovations that improve the specifications of the final good (WTP). Key drivers: Visual Quality, Response Time, Form Factor, Power, Reliability.

2. **Process Innovation ("Raising the Floor"):** Innovations that reduce the effective cost of manufacturing a given unit of quality. Key drivers: Yield, Throughput, Materials, Steps, Repair. Critical tie-breaker: if a structure exists solely to reduce manufacturing steps, it is a Process Innovation.

3. **Mixed Innovation ("The Double Win"):** Innovations that simultaneously achieve a significant product spec improvement AND a significant manufacturing cost reduction.

5-Slot Classification Schema

- **Slot 1 — Type:** Product, Process, or Mixed.
- **Slot 2 — Component:** TFT Array, LC Panel, Backlight Unit, Driver Circuitry, or Module Assembly.
- **Slot 3 — Product Quality Goal:** One of 11 categories (see below), or N/A if Type = Process.
- **Slot 4 — Manufacturing Efficiency Goal:** One of 5 categories (see below), or N/A if Type = Product.
- **Slot 5 — Technological Details:** Natural language summary of the specific technology.

N/A consistency rules enforce logical coherence: Product patents must have Slot 4 = N/A, Process patents must have Slot 3 = N/A, and Mixed patents must have both goals specified.

Product Quality Goals (Slot 3)

1. **Brightness:** Transmittance, aperture ratio, reflectance.
2. **Viewing Angle:** Off-axis color shift/inversion (IPS, MVA, FFS).
3. **Response Time:** Motion blur reduction (overdrive, low-viscosity LC).
4. **Power Efficiency:** Consumption reduction via hardware or logic.
5. **Form Factor:** Compactness, thinness, narrow bezels, module versatility.
6. **Reliability:** Lifetime, ESD protection, thermal stability, seal integrity.
7. **Visual Uniformity:** Flicker, mura, crosstalk, shading, distortion elimination.
8. **Contrast Ratio:** Black levels, light leakage prevention.
9. **Color Accuracy:** Gamma curves, gray scale fidelity, color gamut.
10. **Resolution:** Pixel density, PPI, fine pixels.
11. **Image Quality:** General fallback for picture quality improvements not covered above.

Manufacturing Efficiency Goals (Slot 4)

1. **Yield:** Defect reduction or prevention (design for yield, redundancy).
2. **Steps:** Reduction of photolithography masks or major process steps.
3. **Materials:** Reduction of raw inputs or standardization.
4. **Throughput:** Reduction of cycle time or increased assembly speed.
5. **Repair:** Post-fabrication recovery of defective units.

Output Format

The model returns structured JSON for each patent:

```
{
  "slot_1_type": "Product|Process|Mixed",
  "slot_2_component": "TFT Array|LC Panel|...",
  "slot_3_product_quality_goal": "<goal> or N/A",
  "slot_4_manufacturing_efficiency_goal": "<goal> or N/A",
  "slot_5_technological_details": "<summary>",
  "reasoning": "<1-2 sentence justification>",
  "confidence": <0.0 to 1.0>
}
```

The complete prompt template (167 lines) and the Golden Set of 46 expert-classified examples are available in the replication package.