Collusive Content in Corporate Communications*

Eduardo M. Azevedo† Joseph E. Harrington, Jr.‡ Ioan O. Rusu§

March 5, 2026

Abstract

In industries such as airlines, steel, and truck rentals, firms have coordinated to raise prices or reduce capacities using public announcements such as earnings calls. As this type of coordinating practice is relatively unexplored, this study seeks to learn more about the collusive content in firms' public communications. To do so, we design a large language model (LLM) to look for such content and then apply it to almost 500,000 company transcripts. The LLM is very selective in that only .12% of the transcripts satisfy the criterion for collusive content. The LLM is reasonably effective at rediscovering previously documented episodes and identifies many new episodes. A human audit is conducted of the 301 transcripts with the highest LLM scores. The audit distills and organizes the content to identify systematic ways in which it is facilitating coordinated conduct. Messages fall into two broad bins. First, a firm states that the industry needs to raise prices, reduce supply, or reduce capacity. Second, a firm expresses how it has been less aggressive and calls on competitors to join them, either explicitly or implicitly by saying it is "acting as a leader", commending rival firms who have acted in a similar manner, or criticizing rival firms that have not. A second research objective is to develop a screening tool for competition agencies and plaintiff law firms. Towards that end, the human audit evaluates the performance of the LLM in substituting for human experts. The LLM is capable of identifying a wide variety of collusive content though there is a reasonably high false positive rate. Nevertheless, we explain how it can be an essential component of a screening protocol and show proof of concept with a case study of the Brazilian domestic airline industry.

*We would like to thank Vivian Zhang for research assistance. Replication code available at https://github.com/eduardomazevedo/collusion-llm.

†Department of Business Economics & Public Policy, The Wharton School, University of Pennsylvania, 3733 Spruce Street, Philadelphia, PA 19104. eazevedo@wharton.upenn.edu, https://eduardomazevedo.github.io/.

‡Department of Business Economics & Public Policy, The Wharton School, University of Pennsylvania, 3733 Spruce Street, Philadelphia, PA 19104. harrij@wharton.upenn.edu, https://joeharrington5201922.github.io/.

§School of Economics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104. ir79@drexel.edu.

1 Introduction

It is now well documented (though perhaps not well known) that firms use public communications to collude. Private litigation has delivered compelling evidence of communications through earnings calls and semi-public industry meetings which resulted in firms jointly reducing their capacities in the airlines and steel industries.[1] In addition, plaintiffs have documented public communications to facilitate coordination in the chicken and pork processing industries as well as to supplement private communications for a generic pharmaceuticals cartel.

These public announcements can be quite egregious. In an earnings call, a Vice President for Delta Airlines explicitly called for airlines to coordinate capacity reductions in order to raise fares.

> I think Delta can't do it alone. We have to do it in conjunction with the other carriers because certainly the capacity cuts that we can do on our own, while they will help us, will not remedy the industry's woes. ... And I would say if the industry could achieve a 10% reduction in capacity year-over-year by the fall that we'd be in pretty shape ... [2]

This announcement was one of many made by several airlines calling for "capacity discipline" and commending the industry when it was achieved. The empirical analysis of Aryal et al. (2022) showed that these announcements did cause significant capacity reductions.

There have also been some cases pursued by enforcers. Under Section 5, the Federal Trade Commission issued consent decrees prohibiting certain public communications in the markets for print advertising and one-way truck rentals. The Dutch competition authority obtained a commitment against mobile telecom operators which prohibited them from making certain public announcements relating to their pricing. Most recently, the European Commission has a case against tire manufacturers which "at the heart of the investigation is the novel claim that the tiremakers may have used 'public communications,' in particular earnings calls, as part of a strategy to collude over prices."[3] While there have been few Section 1 prosecutions by the U.S. Department of Justice's Antitrust Division, it is not because such cases cannot succeed.[4] Courts have recognized that "collusive communications can be based upon circumstantial evidence and can occur in speeches at

[1] All of the cases referenced in this and the ensuing paragraph (with the exception of the EU tire manufacturers case) can be found in Harrington (2022).

[2] Delta Air Lines, Inc., Q1 2008 Earnings Conference Call (April 23, 2008). Cited in Harrington (2022), p. 549.

[3] Nicholas Hirst and Jean Comte, "Michelin tells court there's 'no evidence' underpinning EU cartel raids," MLex, March 5, 2025.

[4] To our knowledge, the last case was more than thirty years ago. It involved advance price announcements in the airlines industry (Borenstein (2004)) rather than announcements referencing rival firms' conduct. That case was less egregious than the aforementioned airlines "capacity discipline" episode.

industry conferences, announcements of future prices, statements on earnings calls, and in other public ways."[5]

This study has two objectives. First, it is to advance our understanding of collusion through public communications. This type of collusive practice has only recently been the subject of analysis and, consequently, little is known about it. The second objective is to translate this new knowledge into improved enforcement. There is the concern that the few cases thus far pursued by public and private enforcers may be only the tip of the iceberg.

In order to contribute to both objectives, we first make a methodological contribution by developing and then applying a tool for finding corporate communications containing collusive content. In referring to "collusive content", we mean any content that facilitates coordinated conduct among competitors to reduce competition. Using this tool, we are able to discover many previously unidentified company transcripts containing collusive content. With those transcripts, we distill and organize the associated collusive content in order to better understand how it could affect competition. While we do not explore the effect of this communication on prices or quantities, our study has the ancillary contribution of providing a set of candidate markets for conducting such an empirical analysis. We hope researchers will exploit it. Turning to the enforcement objective, we show how this tool can be used by competition agencies and plaintiff law firms to identify possible cases to prosecute or litigate. In addition, our analysis of the content used in such communications provides the basis for writing guidelines to instruct firms what they should not be saying and what content may be the basis for an investigation.

Achieving these objectives requires discovering communications with collusive content. When collusion involves private communications, the challenge is that firms do their best to keep those communications secret. When it involves public communications, the challenge is quite different. Accessibility is not a problem as the communications are in the public domain. The challenge is instead akin to finding a "needle in a haystack." There is a massive volume of public communications produced by thousands of companies and reviewing it all in search of possible collusive content would seem a fool's errand.

To meet this challenge, we develop a method for reviewing companies' public statements at scale. The approach uses a large language model (LLM). The first step is to properly design a prompt. Guided by some previously documented cases of collusion with public announcements, prompts were experimented with and refined to find one effective at flagging transcripts with collusive content. A scoring process was then constructed by which the LLM assigns a score to each transcript. With the prompt and scoring process in place, around 500,000 transcripts were run through the LLM to look for collusive content. Those transcripts with the highest scores were then reviewed by a human expert. The human audit is integral to both research objectives. First, the audit distills and organizes the content towards understanding the collusive content used by firms. Second, the audit evaluates the LLM's performance and thus assesses its potential as a screening tool for enforcers.

[5] *In re Delta/AirTran Baggage Fee Antitrust Litig.*, 733 F. Supp. 2d 1348, 1360 (N.D. Ga. 2010).

Prior to describing how it contributes to those two objectives, let us first summarize the output of the LLM. To begin, it is very selective as it filters out 99.88% of all transcripts which is consistent with our prior belief that collusive content is rare (though common enough to be an antitrust concern). Critically, the LLM is able to systematically discover companies' announcements with collusive content. In some instances, it rediscovers previously documented communications such as the aforementioned airlines episode. Thus, the LLM was able to find these few collusive "needles" amidst hundreds of thousands of noncollusive "hay."

A concern is that the LLM has been trained on a corpus which includes references to documented cartel episodes, such as those in Aryal et al. (2022) and Harrington (2022) or newspaper stories and legal briefs. In that case, it is only reciting what it learned from "reading." Assuaging that concern, we find that most of the collusive messages found pertain to new episodes. Here are two illustrative examples.

- Kirby Corporation is a leading tank barge operator which transports bulk liquid products in the U.S. river and coastal systems. In its Q4 2017 earnings call, CEO David Grzebinski called on competitors to follow its lead in reducing industry capacity: "We have done our part to rationalize the industry's fleet and improve our profitability going forward, but we need the industry to mirror our actions. If others will follow our lead, an accelerated balance can be restored to the struggling market." He then went on to state how much industry capacity needed to be cut: "I would say 1.5 million to 2 million barrels of capacity needs to come out. We did ... 830,000 barrels ... So if others do their part we could get back in the balance a lot sooner ..."

- During a series of earnings calls, Coca Cola Bottlers Japan conveyed its role as price leader in the beverage market and how raising prices will promote industry profitability. In an earnings call in February 2023, President Calin Dragan stated: "I'm very pleased to see that the entire industry moved to action and acted responsibly last year. So we were able, for the first time, to put the prices up and that's a great thing for our business and for our industry." He went on to say: "This price increase should not be just onetime event ... That's a recipe for failure, not for our business, but for industry. So this, in my opinion, should become a norm in our business here in Japan ..." Later that year, during the Q3 2023 earnings call, executive Costel Mandrea conveyed: "We took leadership in making price decisions in Japanese market, and we will do whatever is needed in order to continue increasing profitability, but also making sure we have a healthy beverage industry in Japan." This point was underscored by President Dragan: "Throughout the years, we were leading in price increases implementation, and in the future, we are not going to be the one which is going to drag back the industry into this because we believe that this is a healthy instrument for the health of the industry."

The LLM was able to discover these messages along with many other transcripts containing collusive content.

The human audit reviews the 301 company transcripts assigned the highest score by the LLM. While there is a rich variety of collusive content, more than 90% of the transcripts can be placed into one of five categories:

1. A firm announces the industry has reduced supply or capacity and references competitors or the industry in a manner supportive of such conduct.

2. A firm announces the industry needs to reduce supply or capacity or to show discipline.

3. A firm announces the industry has shown or will show supply or capacity or price discipline.

4. A firm announces the industry needs to raise prices.

5. A firm announces it is being or will be less aggressive in pricing and references competitors' conduct in a manner encouraging similar conduct.

This study informs us of the manner in which public announcements facilitate coordinated conduct. Coordination is an understudied area compared to research on how firms design the collusive agreement (i.e., collusive strategies) and achieve cartel stability (i.e., satisfy equilibrium conditions).[6] Especially when the method of coordination does not involve direct, express, and private communication (i.e., the proverbial "smoke-filled" room), it is important to understand how firms communicate as that can affect the form collusion takes (e.g., firms may raise prices but without a market allocation scheme) as well as the likelihood of collusion (e.g., whether firms achieve mutual understanding to raise prices).[7]

The study's second objective is to contribute to enforcement by developing a screening tool for collusion involving public communciations. The LLM's ability to identify a diverse array of collusive content is impressive. At the same time, the false positive rate is 64% (where a false positive means the LLM says there is collusive content but the human expert disagrees). However, if only a small percentage of company transcripts genuinely have collusive content - which seems reasonable - then even a low error rate will produce many false positives. Thus, a high false positive rate may be inevitable. What is crucial for enforcement is that the LLM identifies collusive content that can be the basis for an investigation. Our takeaway is that an LLM, when complemented with human oversight, can be part of an effective screening program. To show proof of concept, we provide a

[6] In the language of game theory, research is almost entirely focused on understanding the collusive equilibrium upon which firms settle. Coordination is about how firms move from a competitive (static Nash) equilibrium to a collusive (repeated game) equilibrium.

[7] Our study focuses on a particular class of public announcements. Other less explicit coordinating practices include advance price announcements (Borenstein (2004), Chen et al. (2017)), coded messages (Cramton and Schwartz (2000), Harrington (2025)), and price signaling (Byrne and de Roos (2019), Byrne et al. (2025)).

case study of the Brazilian domestic airline industry. The LLM identifies suspicious content and subsequent examination reveals convincing evidence of two airlines effectively communicating to coordinate a reduction in their capacities.

This second contribution adds to the growing literature on cartel screening. Generally, screening involves looking for collusion in data sets comprising prices or bids (in the context of a procurement auction). This could mean testing for collusive markers (i.e., patterns more consistent with collusion than competition) or for structural breaks associated with cartel birth or death.[8] In contrast, our method works on text rather than numerical data and fits into a body of work on "text as data" (Gentzkow et al. (2019)).

Literature Review

Most relevant to our study is the recent literature on the use of public disclosures to facilitate collusion. Exploring the choice of firms to collude through private or public communications, Bourveau et al. (2020) find that as cartel enforcement against explicit collusion increases - as measured by the rise in leniency programs - firms reveal more proprietary information through financial disclosures including earnings calls. They argue that this information sharing is conducive to tacit collusion. Kepler (2021) also looks at the trade-off between public and private information sharing in the context of firms' coordinating prices or outputs. He finds that when firms form strategic alliances - which gives them the opportunity to privately exchange information - firms disclose less about their expected future business conditions and production decisions in their public announcements. Pawliczek et al. (2022) test whether firms having stronger incentives to collude - and thereby are more likely to be trying to collude - results in more public disclosures. The source of variation in the proclivity to collude is the degree of common ownership across firms. As common ownership increases, firms disclose more about business conditions and strategies in their earnings calls.

The next three studies more directly examine the relationship between public announcements and collusive conduct. Harrington (2022) collects and categorizes all known cases in which there is evidence that firms coordinated using public announcements which reference rival firms. A typology of announcements is created which is reviewed in Section 4. Aryal et al. (2022) and Sheng and Vukina (2024) measure the effect of public announcements on firms' conduct. For the airlines industry, Aryal et al. (2022) find that when all airlines on a route have "capacity discipline" or related content in their earnings calls then their capacities (as measured by seats) decline by around 2% in the ensuing quarter. This decrease is more than half of the average inter-quarter variation. For the broiler chicken industry, Sheng and Vukina (2024) examine the effect on various output measures of the frequency of firms using the words "cut, balance, constrain, discipline, reduction, or adjustment" in their earning calls. A 1% increase in their frequency is associated with a 1-1.5% decrease in output depending on the measure.

The preceding three studies focused on previously documented cases of collusion.

[8] Harrington (2008), Harrington and Imhof (2022), and *Detecting Cartels for Ex Officio Investigations* (2024) offer surveys of cartel screening. For some recent research, see Huber and Imhof (2019), Kawai and Nakabayashi (2022), Chassang et al. (2022), Kawai et al. (2023), Silveira et al. (2023), and Aaltio et al. (2025).

Our study's objectives require finding new episodes in which public announcements facilitate coordinated conduct. The most closely related study is Duso et al. (2025) which is the first to apply Natural Language Processing to a large corpus of company public announcements in search of collusion. Rather than use an LLM, they use a dictionary approach as in Aryal et al. (2022) and Sheng and Vukina (2024). Bigrams - or pairs of words - are defined to capture a firm referring to its strategy or to its competitors; e.g. "competitor should" and "raise price". The score for a firm in a quarter is the number of times those bigrams appear in its public announcements for that quarter. The study provides examples showing how this screening approach can identify public announcements with collusive content. Our study differs in terms of methods and objectives and, more specifically, the use of an LLM and the conducting of a human audit. The human audit allows us to evaluate the performance of our LLM-based screen and to advance our understanding of the content used by firms to coordinate in reducing competition.

Roadmap

The data used in our analysis along with the development process for the LLM screen are described in Section 2. The output from the large-scale LLM analysis is summarized in Section 3. In Section 4, the procedure for the human audit is described and the performance of the LLM is evaluated. The content identified by the LLM is distilled and organized in Section 5. Section 6 discusses the role of an LLM in a screening protocol. Finally, a case study of the Brazilian domestic airline industry is provided in Section 7.

2 Data and Methods

2.1 Data

Our transcript text data comes from the Capital IQ transcripts database (currently branded as S&P Global Machine Readable Transcripts). The dataset includes earnings calls, investor day presentations, industry conferences, analyst meetings, and other public corporate communications.

The transcripts span from 2002-02-14 to 2025-04-11, covering 18,694 companies. Our dataset comprises 498,602 transcripts. The average call lasts 48 minutes 56 seconds. In total, it would take 43 years of uninterrupted audio to listen to all calls. Capital IQ provides verbatim transcriptions of these calls, including remarks by company executives and the subsequent question-and-answer sessions with analysts and investors. Speakers are identified and tracked across the transcripts.

The data includes company identifiers, call dates, and call types. We merge Compustat data at the company-year level to obtain basic information like sector and market capitalization.

We use expert ratings of collusive and non-collusive transcripts from Aryal et al. (2022).[9] They collected earnings calls from 11 airlines from 2002 to 2016, a period dur-

[9] https://doi.org/10.5281/zenodo.5664420

Table 1: Summary statistics for dataset characteristics

Variable	Mean	Median	Min	Max	*N*
Market Value (mil USD)	15,466.7	1,888.9	0.024	3,761,715	291,558
Employees (thousands)	20.6	3.5	0.001	2,300	418,626
Audio Length (seconds)	2,936.8	2,657.0	120	17,998	466,433
Transcript Year	2017.7	2018	2002	2025	498,602

Notes: Reports mean, median, min, max, and non-missing *N* for continuous variables describing the dataset characteristics.

ing which major airlines allegedly tried to enforce "capacity discipline" through public communications. They rated the transcripts as collusive or non-collusive with a keyword search and manual review. They only coded as collusive the transcripts where all three authors agreed that the transcript was collusive. Their data includes 503 transcripts accessed through the LexisNexis database and the investment news website Seeking Alpha, with about 33% being coded as collusive. We use 298 of these transcripts that are available in the Capital IQ dataset.

2.2 Methods

Our analysis pipeline is illustrated in Figure 1. We take the following steps.

Construction of human-labeled benchmark sample

We constructed a human-labeled benchmark sample using two sources. The first is the Aryal et al. (2022) transcripts, with their binary assessment of collusion. The second are transcripts from U-Haul's parent company, Amerco. Some of these transcripts were previously discussed and considered collusive by Harrington (2022). Citing these communications, the FTC charged U-Haul with an "invitation to collude" in *In the Matter of U-Haul International, Inc. and AMERCO* (2010), resulting in a settlement. We reviewed these transcripts ourselves and provided collusion severity scores on a 0-100 scale. To obtain a binary classification for the U-Haul sample, we set a threshold value of 50. This leaves us with a benchmark sample of 389 transcripts.

Development and evaluation of prompts and LLM models

With the benchmark sample in hand, we developed and evaluated prompts and experimented with different LLM models (see Appendix A for prompt comparison).[10] We ultimately selected *gpt-4o-mini* (temperature=1.0) and the following zero-shot prompt:

> *You're an expert in detecting subtle collusive intent in company earnings calls. You know firms can try to avoid scrutiny if they express intentions to reduce supply or*

[10]Recent work by Asirvatham et al. (2026) supports our approach in prompting, variable extraction, and benchmarking through a more generalized package that uses OpenAI LLMs on a selection of tasks. They also highlight using such LLMs as "measurement tools" and showcase ways to improve reliability and validity of LLM output.

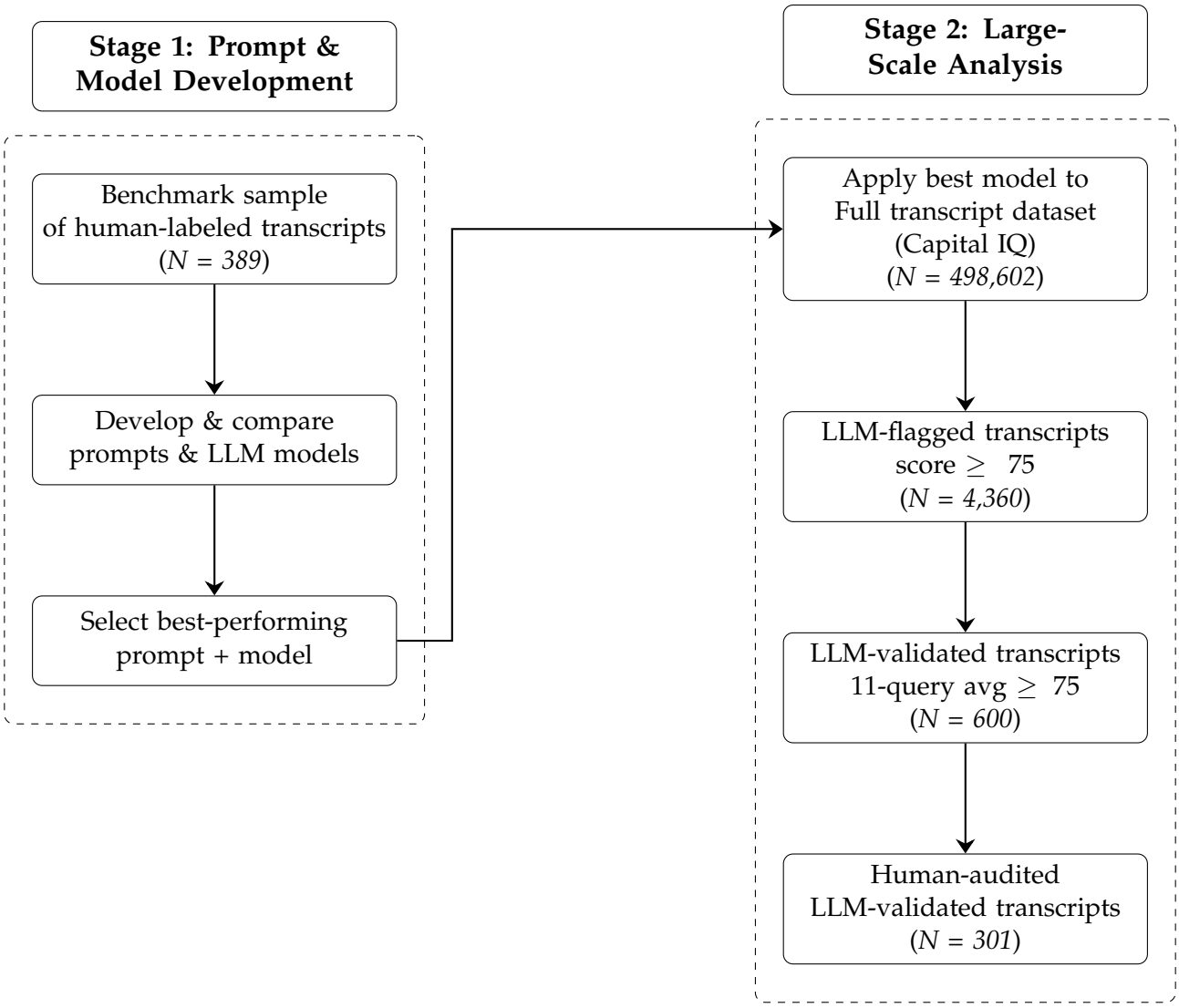

Figure 1: Analysis Workflow

Notes: This figure illustrates the workflow for detecting collusive corporate communications. Stage 1 develops and evaluates prompts and models using the human-labeled benchmark dataset through cross-validation to select the best-performing combination. Stage 2 applies the chosen model at scale to the full Capital IQ transcript dataset (498,602 transcripts), refines high-scoring transcripts with multiple queries using a 75-point threshold, and conducts human audits of the top-scoring results to validate model performance.

> *increase prices, or they can imply it would be good if industry capacity were limited or prices were kept high. Analyze the following transcript of a public company communication and identify any signs that the company is trying to coordinate with competitors to limit capacity or keep prices high. Provide a score from 0 to 100, where 0 means no evidence of collusive intent and 100 means strong evidence of collusive in-*

tent. Additionally, provide a very brief explanation of your assessment. Also provide a list of excerpts that support your assessment.

This prompt was crafted to mention the broad types of collusive actions (i.e. maintaining high prices, limiting capacity), and specify clearly that we target a company's intention to coordinate on such actions.

Large-scale analysis of Capital IQ transcripts

We applied the best-performing prompt and model to the full Capital IQ transcript dataset (N=498,602). We define a transcript as **LLM-flagged** if it has an initial collusion score of at least 75. These are transcripts where the initial LLM assessment was that there is significant evidence of collusive intent. This flagged 4,360 transcripts.

LLM Validation of LLM-flagged transcripts

Because the LLM we use generates non-deterministic outputs for the same given transcript/prompt pair, a single evaluation of a transcript may reflect idiosyncratic sampling noise rather than a stable assessment. To address this issue and to produce a more robust output, we ran an additional ten repeated queries on the LLM-flagged transcripts. We then define as **LLM-validated** transcripts those whose average score remained at least 75 after the validation. This leaves us with 600 transcripts.

Human audit of LLM-validated transcripts

Finally, we manually audited the top 301 LLM-validated transcripts. The audit is crucial for both of the study's objectives: understanding the content used to facilitate coordinated conduct to reduce competition and evaluating the performance of an LLM as a screen for enforcers. The audit procedure is described in section 4.

2.3 Rationale for Using an LLM

Text analysis in economics has traditionally proceeded by mapping documents into structured numerical representations, such as dictionaries that count curated terms and phrases, bag-of-words or TF–IDF features that count word occurrences, and topic or factor models that compress word counts into a small number of latent dimensions (Gentzkow et al. (2019); Blei et al. (2003)). These approaches have been used successfully in a wide range of applications, including dictionary-based policy uncertainty indices (Baker et al. (2016)), transcript-based political risk measures (Hassan et al. (2019)), and text-based measures of industry similarity (Hoberg and Phillips (2016)). A common feature of these methods is that they summarize text through word frequencies or low-dimensional transformations of those frequencies, after which standard statistical tools are applied.

Recent advances in natural language processing instead rely on large language models (LLMs), which process text as sequences and form context-dependent representations of words and phrases. Modern LLMs are typically built on the transformer architecture (Vaswani et al. (2017)), which uses an attention mechanism to allow each word in a document to be interpreted in light of the surrounding text, including both nearby and more distant context. In addition, many contemporary models are instruction-tuned to follow natural-language task descriptions rather than merely predict the next word in

a sequence (Ouyang et al. (2022)). In our setting, we use such an instruction-following LLM in a prompt-based manner, treating it as a document annotator that receives a transcript and a task description and returns a structured assessment. This differs from traditional feature-based pipelines in that the mapping from text to output is generated directly by the model in response to the prompt, rather than through pre-specified word lists or estimated low-dimensional representations. We do not claim that this approach dominates alternative text-analysis techniques; rather, it reflects the view, supported by a growing literature on the use of LLMs in economics (Korinek (2023); Ludwig et al. (2025)), that context-sensitive, instruction-following models provide a natural additional tool for large-scale analysis of unstructured communications.

3 Results from Large-Scale LLM Analysis

3.1 Descriptive Statistics of LLM Scores

We begin with the results from the large-scale LLM analysis, prior to the human audit. Table 2 shows the summary statistics of the classification flags and LLM scores.

The first obvious feature of the data is that very few transcripts are flagged as collusive by the LLM. Only 0.9% meet the threshold of 75 collusive score in the original query. This is an important sanity check because presumably public collusive communications are rare. Moreover, the LLM is confident in its judgement that most transcripts are not collusive. We see this both from the low average score of 20.8, and from the histogram in Figure 2 which is concentrated on the left.

We now turn to the validation step, where each flagged transcript is re-queried 10 times. Out of all the flagged transcripts, only 14% pass the validation threshold of an average score of at least 75. The validation step leaves us with a sample of 600 LLM-validated transcripts.

This winner's curse can be seen transparently in the data. Figure 3 shows that the distribution of mean scores has the bulk of its values below 75. And Figure 4 shows that the mean scores are often below the original scores. Thus, most LLM-flagged observations are due to the LLM randomly producing a high score. At the same time, both figures show that transcripts with a high original score tend to have a high mean score after validation. This suggests that the LLM is capturing some signal. We will confirm this below, as we turn to the benchmark sample, human audit, and detailed reading of some transcripts.

3.2 LLM Performance in the Benchmark Sample

We now measure the performance of the LLMs on the benchmark sample. We are interested both in the LLM's ability to discover collusive communications, and its ability to avoid flagging non-collusive communications as collusive.

We first count how many of the collusive communications in our benchmark human-rated dataset are rediscovered by the LLMs. In our benchmark dataset, we have 135

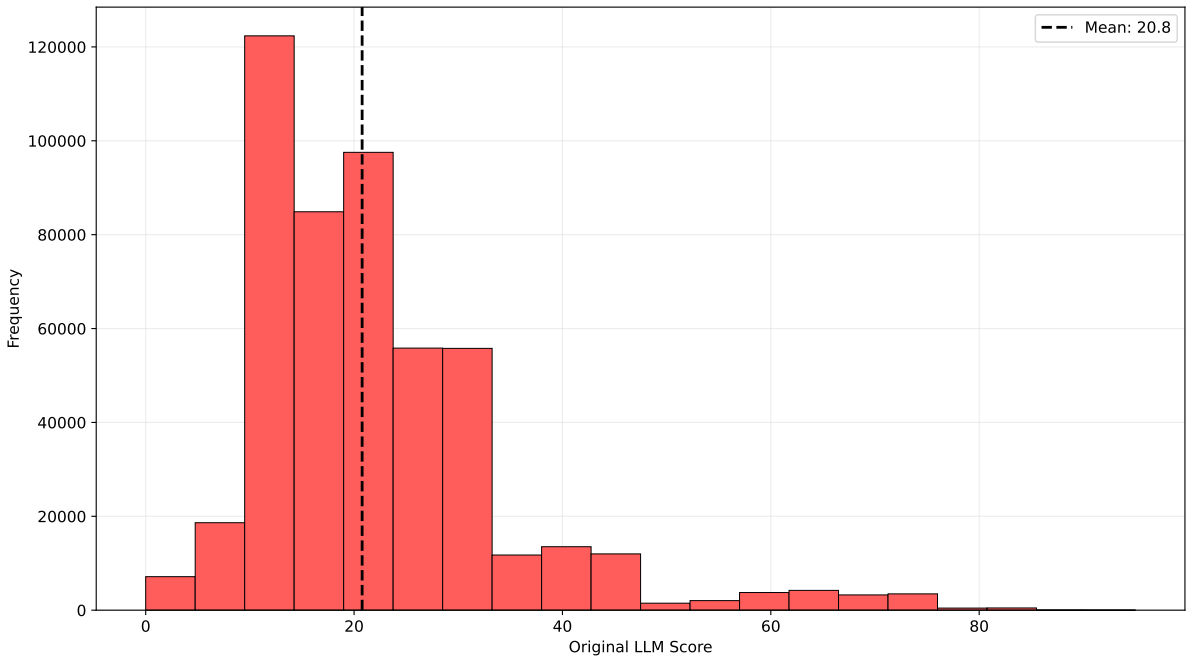

Figure 2: Distribution of original LLM scores for the entire sample.

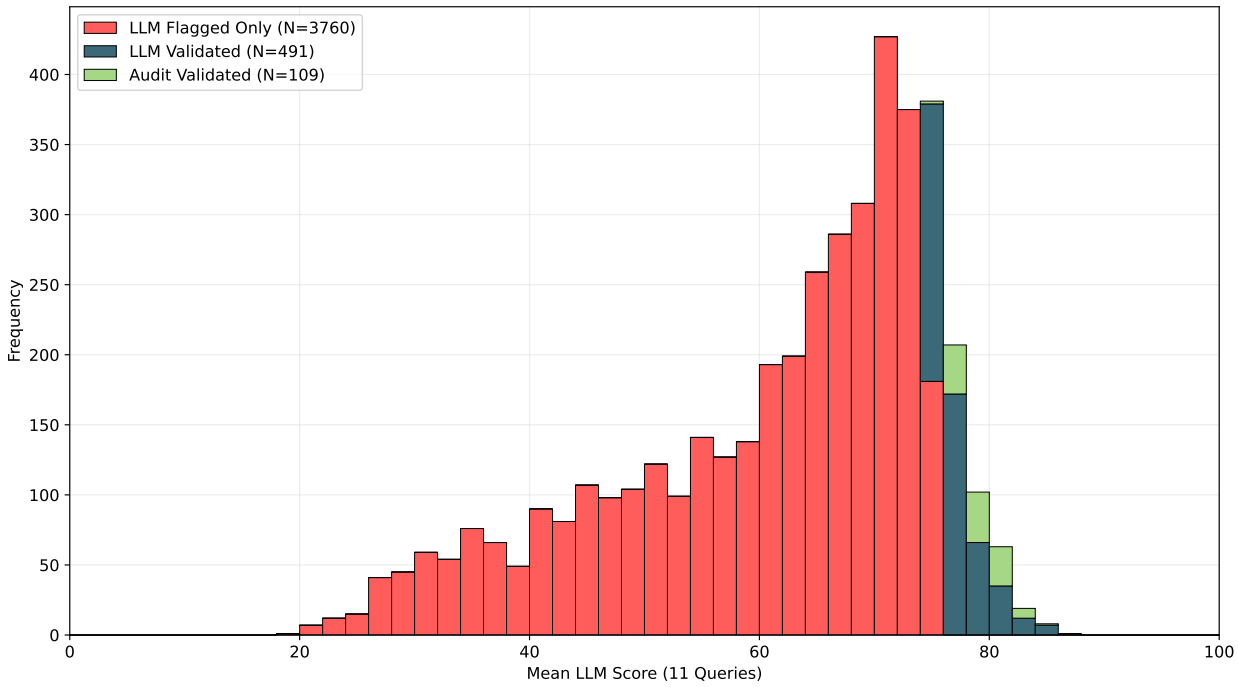

Figure 3: Distribution of mean LLM scores (with 10 repetitions) for LLM flagged transcripts.

Table 2: Summary statistics for classification results

Panel A: Boolean classification flags

Variable	*N*	Count (True)	Percent (True)
LLM Flagged Collusive	498,602	4,360	0.9%
LLM Validation Flag	4,360	600	13.8%
Human Audit Sample	498,602	301	0.1%
Human Audit Flagged Collusive	301	109	36.2%
In Benchmark Sample	498,602	389	0.1%
Human Benchmark Flagged Collusive	389	135	34.7%

Panel B: LLM score statistics by sample

Sample	Original Score		Mean Score (11 Queries)	
	Mean (SD)	*N*	Mean (SD)	*N*
Entire Sample	20.77 (13.00)	498,602	—	–
LLM Flagged	76.66 (3.42)	4,360	61.75 (13.79)	4,360
LLM Validated	78.54 (4.47)	600	77.47 (2.19)	600
Audit Validated	80.05 (4.59)	109	79.10 (1.72)	109

Notes: Panel A reports non-missing *N*, count of TRUE values, and percent TRUE for key boolean classification flags. Panel B reports mean (standard deviation) and sample size *N* for LLM scores across different samples. Original Score is from the initial query. Mean Score (11 Queries) is the average across 11 queries (original query plus 10 follow-up queries), and is only available for LLM flagged transcripts. "—" indicates the score is not available for that sample.

collusive communications that human experts rated as collusive. Out of these, 42 (31%) were flagged in the large-scale LLM queries, and 17 (13%) were validated as collusive. While this is far below 100%, it is also far above zero. This suggests that LLMs are able to discover a significant percentage of collusive communications in a large sample, even if imperfectly.

To examine the rate of false positives, we now count how many of the non-collusive communications in our benchmark dataset were flagged as collusive by the LLMs. In our benchmark dataset, we have 254 transcripts that human experts rated as not collusive. Out of these, 8 (3.1%) were flagged in the large-scale LLM queries, and 1 (0.4%) was validated as collusive.

These numbers show that there is a significant signal in the LLM output. We can see this in the stark difference in the rate at which the collusive and non-collusive communications are flagged by the LLM. Transcripts that humans rated as collusive were 36 times more likely to be validated as collusive by the LLM. This suggests that while LLMs are imperfect, they meaningfully distinguish between collusive and non-collusive communi-

cations.

3.3 Correlates of Collusive Content

We now examine the correlates of collusive communication and consider three types: firm sector, firm size, and call year. LLM-flagged is used as our binary indicator of collusive communication. This has the issue of being a noisy proxy for collusive communication. But the other measures are also imperfect, and LLM-flagged has the advantage of a larger sample size.

The clearest pattern is the heterogeneity by sector, displayed in Figures 5 and 6. We split the sample into top level GICS sector definitions. There is considerable heterogeneity around the average rate of collusive communications of 0.9%. The sector with the lowest rate of collusive communications is health care (0.2%), and the highest is materials (3.0%). Encompassing many commodities, the high prevalence of collusive content for the materials sector is consistent with the empirical evidence on cartels.

Based on our prior knowledge of cartels, we identified the three-digit sectors most prone to collusion. They encompass 18 sectors listed in Figure 6 (including the legend). Figure 6 shows that all but two sectors have rates of collusive content higher than the average, and often significantly so. Pulp Mills, Flat Glass, and Paper Mills have a rate well ranging from 11 to 14 times the average rate. Concrete, Gypsum, Plaster Products and Cement, Hydraulic have a rate 6-7 times the average rate. That Heavy Construction is below average is likely due to collusion tending to occur through bidding rings at procurement auctions. This type of collusion is not facilitated by public communications as in the other high collusion segments.

The patterns by market value decile are less clear (Figure 7). The largest firms, except in the top decile, have a slightly higher than average probability of collusive communication. While the differences are statistically significant, they are small.

Likewise, Figure 8 suggests a mild decrease in collusive communication over time. While we see statistically significant differences across these years, these are relatively weak and noisy patterns.

Summary

The output of the LLM is highly selective. An initial query with a threshold score of 75 weeds out more than 99% of company transcripts. However, there is a randomness in the LLM's output which makes it critical to run multiple queries. Doing so another ten times and using the mean score weeds out another 86%. This leaves us with 0.12% of the corpus. Critically, the LLM is able to rediscover a substantive number of the collusive messages from the benchmark sample and almost entirely excludes the noncollusive messages for a very low false positive rate. Finally, it is finding much more collusive content in sectors known to be prone for collusion.

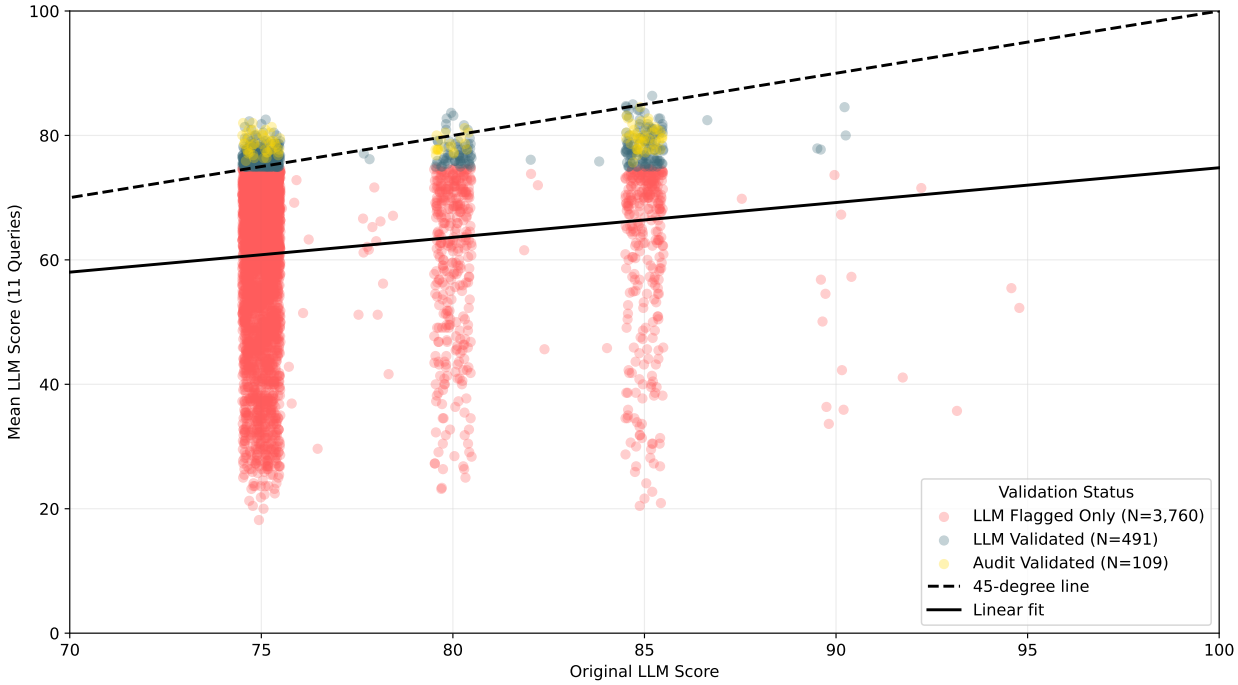

Figure 4: Original LLM Score vs. Mean LLM Score after 10 repetitions.

Notes: Points have horizontal jitter (± 0.5) to improve visibility.

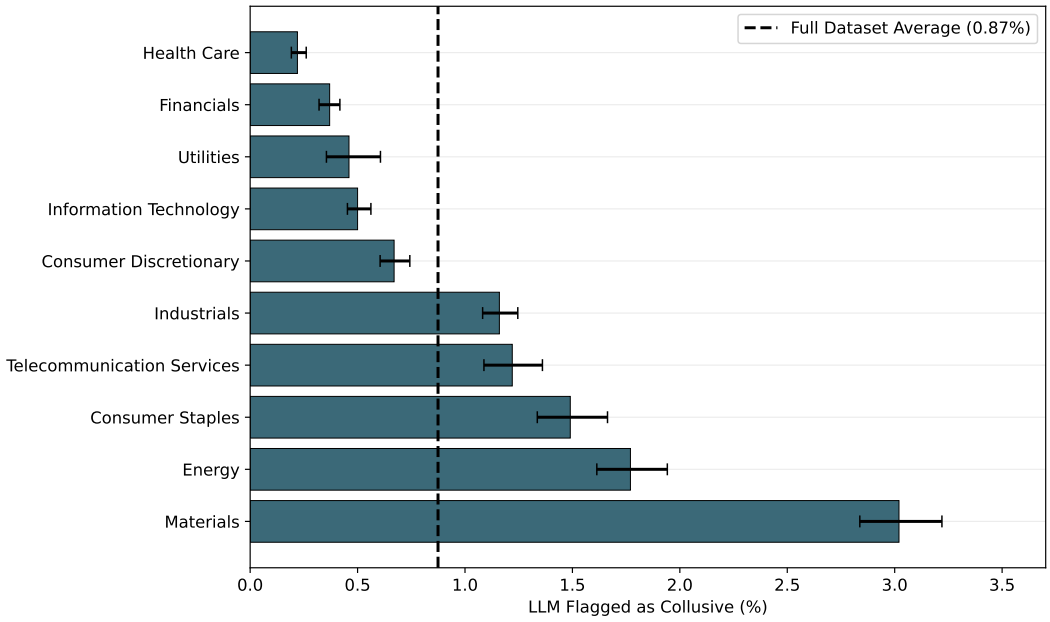

Figure 5: LLM Flagged Rate by Sector

Notes: Fraction of LLM-flagged collusive communications across industry sectors (GICS) represented in the full transcript dataset, excluding GICS 60–Real Estate. Error bars represent 95% confidence intervals.

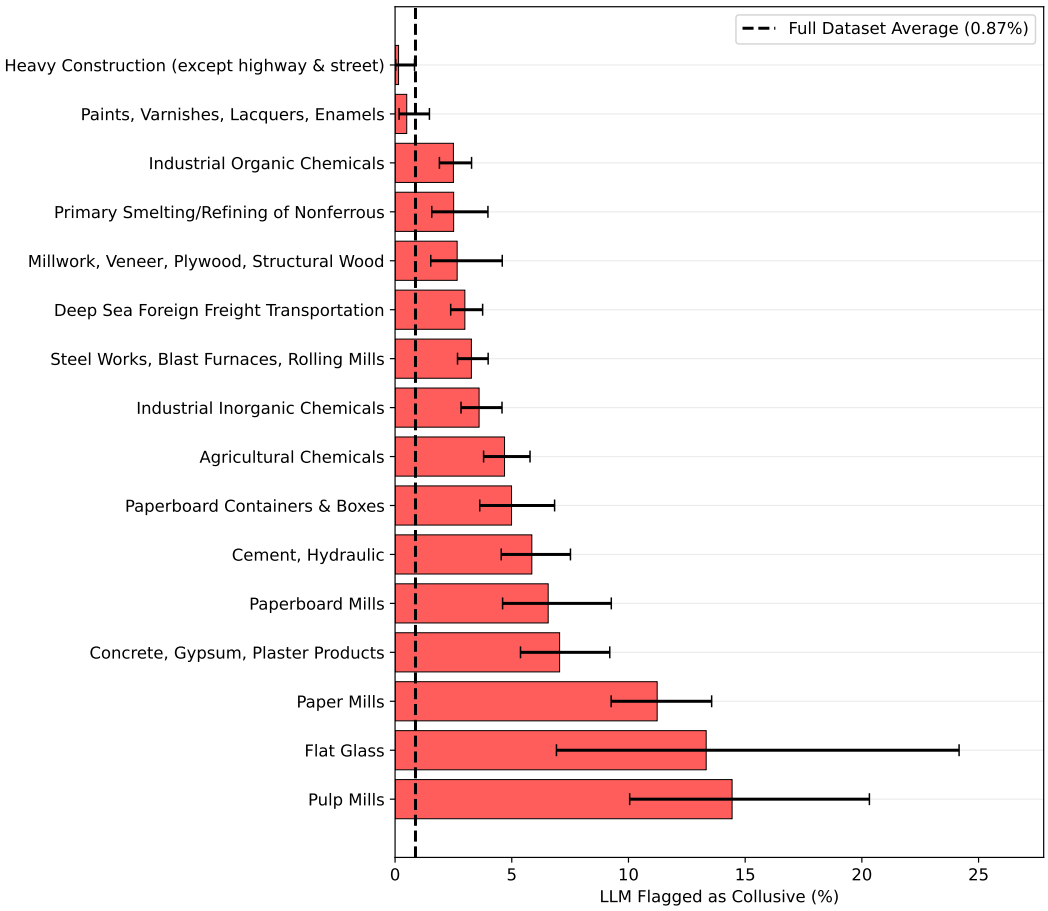

Figure 6: LLM Flagged Rate by High Collusion Segments

Notes: Fraction of LLM-flagged collusive communications across segments that are prone to collusion, excluding Deep Sea Domestic Freight Transporation and Highway & Street Construction, which don't appear in our transcript dataset. These segments were selected based on theoretical predictions of collusion risk (e.g., homogeneous products, high transport costs, capacity constraints) and historical cartel cases. Error bars represent 95% confidence intervals.

4 Audit

4.1 Description of Audit Procedure

This study seeks to contribute to our understanding of the use of public announcements to collude and to developing a screening tool for detecting such public announcements. The human audit serves both those objectives.[11] First, it distills and organizes collusive content towards understanding the type of content that is being used to coordinate firms' conduct. Second, it evaluates the performance of the LLM and thereby assesses how well

[11]The auditor is one of the paper's authors, Joseph Harrington.

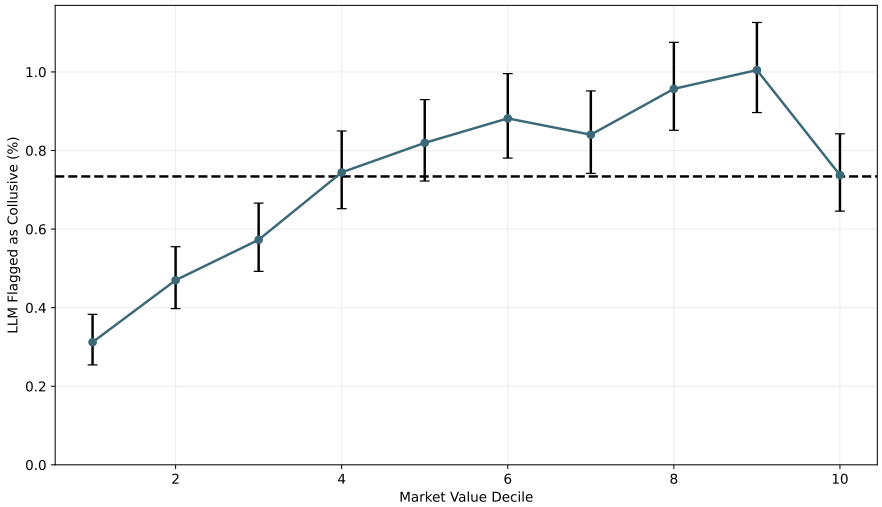

Figure 7: LLM Flagged Rate by Market Value Decile

Notes: This figure shows the fraction of communications tagged as collusive by the LLM across market value deciles, representing firm size. The chart displays how collusive communication detection varies with company size. Data includes 291,558 transcripts in the size analysis. Error bars represent 95% confidence intervals.

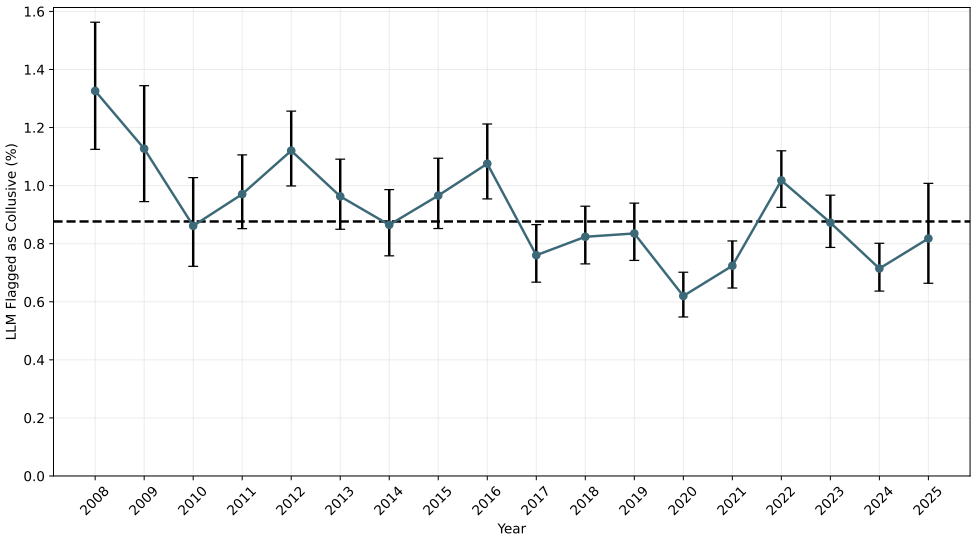

Figure 8: LLM Flagged Rate by Year

Notes: This figure shows the fraction of communications tagged as collusive by the LLM over time. The chart displays temporal trends in collusive communication detection across the sample period. Data includes 498,602 transcripts in the call year analysis. Error bars represent 95% confidence intervals.

it does in substituting for human experts. In this section, the procedure for the human audit is described and the performance of the LLM is evaluated. In Section 5, the collusive content is analyzed and organized to deliver some takeaways about how firms are seeking to reduce competition.

Balancing the value of reviewing more transcripts against the time it takes to review them, we decided to review the top 300 transcripts - which makes up one-half of the LLM-validated transcripts - subject to including all transcripts with the same score. This resulted in 301 transcripts which comprise all those with a score of at least 76.92. The scores ranged from 76.92 to 85.00 and are summarized in Table 3 and Figure 9.[12]

Table 3: Human audit summary statistics

Sample	N	Score min	Score max	Score mean (SD)
Human audit sample	301	76.92	85.00	79.05 (1.83)

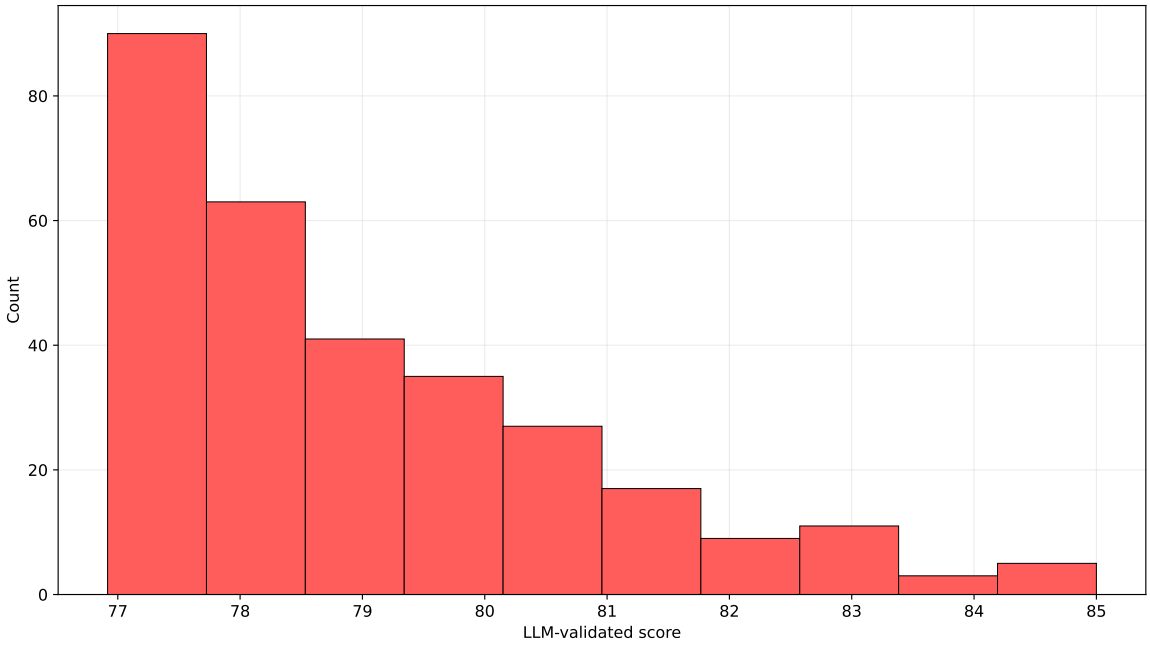

Figure 9: Distribution of LLM-validated scores in the human audit sample

For each of these transcripts, the human auditor read the excerpts provided by the LLM.[13] In order to make for a more independent judgment, the human auditor did not

[12] Actually, there are 302 transcripts satisfying this criterion but one was removed as the content described *past* collusion. It was an earnings call from Loblaw Companies which was a member of the Canadian bread cartel. Notably, the transcript received a score of 86.25 which was the highest score in the sample.

[13] Quoting from the prompt, the LLM was instructed to "provide a list of excerpts that support your

read the LLM's explanation for its score.[14] The human auditor was aware that the LLM had given the transcript a high score and thus knew the LLM had rated the content as collusive. It is possible this knowledge could bias their evaluation though we have no reason to think it did.[15]

The human auditor's evaluation was guided by the typology in Harrington (2022). What potentially distinguishes collusion-facilitating content is that a firm makes reference to the conduct of rival firms or the industry at large. Within that class of public announcements, three categories are identified in Harrington (2022). First, a firm describes how its future conduct depends on a rival firm's conduct. A firm may express it is acting as a leader with the implicit message that rival firms are to follow, or as a follower which then invites a rival firm to lead. Second, a firm prescribes how rival firms or the industry should behave in the future. This could mean explicitly recommending conduct in terms of such variables as prices and capacities. Or a firm could commend competitors or the industry for past conduct and thereby implicitly recommend continuation of that conduct or criticize past conduct and thereby implicitly recommend discontinuation of that conduct. Third, a firm describes how rival firms or the industry will behave in the future. This forecast could be an invitation to firms to act consistent with that forecast.

If the human auditor found the transcript's excerpts contain collusive content then it was classified as a "true positive" (and labelled T), and if the human auditor did not find the transcript contains collusive content then it was classified as a "false positive" (and labelled F). If T then the human auditor summarized the collusive content. If F, the human auditor summarized the content that might have led the LLM to give the transcript a high score. In some cases, there is no relevant content in which case it is labelled "no content". The LLM's output along with the human auditor's evaluation and summary are provided in Appendix B.[16]

4.2 Evaluation of LLM Performance

Of the 301 transcripts with the highest LLM-validated score, the human auditor rated 109 as containing collusive content and 192 as lacking collusive content for a true positive

assessment." Reading only the excerpts and not the entire transcript was done as a time-saving device. It could downward bias the true positive rate if the LLM properly identifies a transcript as containing collusive content but does not always include sufficient relevant material in the excerpts to allow a human auditor to draw the same conclusion. Another reason for focusing on the excerpts is that a competition agency's protocol in using an LLM could involve reading only the excerpts in order to save on the agency's human resources. In that case, the LLM's efficacy in extracting passages to put in excerpts is relevant to its performance.

[14] Quoting from the prompt, the LLM was instructed to "provide a very brief explanation of your assessment."

[15] All information used by the human auditor is available in Appendix B so that the reader may make their own judgment.

[16] Of course, the human auditor may be mistaken in their judgment - in the sense that the content did not result in firms coordinating their conduct to reduce competition - but the goal is not to identify actual episodes of collusion but to determine if an LLM can replace human experts.

rate of 36.2%. Table 4 reports the true positive rate after partitioning the sample into three roughly equal-sized subsamples (while respecting score breaks). The true positive rate is decreasing in the score with a true positive rate of 41.5% for the 106 transcripts in the highest bin and 31.9% for the 94 transcripts in the lowest bin.

Table 4: LLM Performance Statistics

Sample	Scores	Number	True positive	False positive	True positive rate
Full sample	76.92-85.00	301	109	192	36.2%
Top bin	79.58-85.00	106	44	62	41.5%
Middle bin	77.92-79.38	101	35	66	34.7%
Bottom bin	76.92-77.85	94	30	64	31.9%

There are 71 distinct companies in these 109 true positive transcripts. Table C.1 in Appendix C lists all companies along with their primary market. The four companies with the largest number of transcripts are GOL Linhas Aéreas Inteligentes (7 transcripts) which is a domestic airline in Brazil, Micron Technology (6) which is a manufacturer of computer memory and data storage, Norske Skogindustrier (4) which is a pulp and paper manufacturer, and Coca-Cola Bottlers Japan (4) which was mentioned in the Introduction. We will take a closer look at GOL Linhas Aéreas Inteligentes in our case study.

21 of the 109 transcripts are from previously documented episodes of companies making public announcements with collusive content, including 15 associated with the airlines case. 10 of those 15 airlines transcripts were from companies in the Aryal et al. (2022) study: US Airways (3 transcripts), American Airlines (2), Delta Airlines (2), United Airlines (2), and AirTran (1); and then there were five transcripts from Air France-KLM (3) and Frontier Airlines (2). The other previously documented episodes are: broiler chicken (Pilgrim's Pride), pork (Smithfield Foods, WH Group), and generic drugs (Teva Pharmaceuticals). In some of these cases – generic drugs for sure and possibly broiler chicken and pork – collusion also involved the more standard private communications. The LLM flagging them is a reminder that screening public communications for collusive content could lead to investigations that uncover the most compelling evidence of illegal collusion: direct, express, and private communications between competitors.

In the next section, we will analyze the content of the true positives in order to better understand how firms are faciliating coordinated conduct. Here, let us comment on the false positives to understand how the LLM is getting it wrong. Broadly speaking, a transcript was designated as having collusive content (and thus being a "true positive") when a company makes reference to some form of less aggressive conduct (such as lower capacities and higher prices) and connects it to competitors or the industry. In examining the 192 false positives, a content analysis finds that 163 of them refer to the firm's own conduct or the state of market competition though did not connect it to the conduct of competitors or the industry and thereby fell short of facilitating coordinated conduct.[17] 67 transcripts referred to a firm reducing, restricting, or managing its supply or capac-

[17]Details from the content analysis can be found in Appendix E.

ity in a manner consistent with higher prices or avoiding lower prices. However, it did not say or suggest that this conduct is something the industry should or will or needs to do. 48 of the transcripts refer to higher prices - whether noting having raised prices or expressing a need to raise prices - or to pricing power. Such a statement might be made in the context of reducing supply. In 14 transcripts, the firm refers to industry consolidation by noting that it has been consolidating, needs to consolidate, or will consolidate. In some instances, the firm attributes consolidation to too much supply or capacity or too many firms. In sum, while the LLM identified instances in which a firm's message was consistent with competing less aggressively, it did not have the content that encourages similar conduct by competitors. The other 29 false positive transcripts lacked any relevant content in which case it may be appropriate to classify them as hallucinations. With that definition, the hallucination rate is 9.6% (29 out of 301).[18]

In concluding, let us note that that a high false positive rate of 64% is not surprising if it is rare for a firm to have collusive content in its public announcements. Even if the error rate (i.e., the likelihood that a transcript without collusive content is given a high score by the LLM) is low, a high fraction of transcripts with a high LLM score will not have collusive content when sufficently many transcripts lack collusive content. Consistent with this hypothesis, the error rate was a very low 0.4% for the benchmark sample.[19] Overall, the LLM has performed well in being able to find many "collusive" needles in the haystack even if there is some hay in the basket of needles.

5 Synthesis of Collusive Content

5.1 Overview

As described in the preceding section, the human audit reviewed the 301 transcripts with the highest LLM-validated scores and found 109 of them contained collusive content. We now want to scrutinize the LLM-generated excerpts from those 109 transcripts in order to understand how that content is facilitating coordinated conduct to reduce competition. Though the content is diverse and spans a wide set of markets, we find that more than 90% of the content falls into one of five categories:[20]

[18]Of course, these statements are based on a reading of the excerpts and it is possible there is more incriminating material elsewhere in the transcript. However, the LLM was instructed to provide content supporting their assessment so either they did a poor job with that task or the score is not justified by the content.

[19]With the audit sample, we are measuring the likelihood that the transcript does not have collusive content conditional on it being assigned a high score by the LLM. With the benchmark sample, we are measuring the likelihood of the LLM assigning a high score conditional on the transcript not having collusive content.

[20]Some excerpts have content spanning more than one category, which will be apparent in some of the quoted passages. When the entire transcript is read, there are likely to be even more cases in which content encompassed multiple categories. For this reason, the number assigned to each category is not meant to be definitive but only to give a rough sense of its frequency.

1. A firm announces the industry has reduced supply or capacity and references competitors or the industry in a manner supportive of such conduct. (40 transcripts)
2. A firm announces the industry needs to reduce supply or capacity or to show discipline. (30 transcripts)
3. A firm announces the industry has shown or will show supply or capacity or price discipline. (12 transcripts)
4. A firm announces the industry needs to raise prices. (12 transcripts)
5. A firm announces it is being or will be less aggressive in pricing and references competitors' conduct in a manner encouraging similar conduct. (8 transcripts)

5.2 Category 1: The industry has reduced supply or capacity and references competitors or the industry in a manner supportive of such conduct

The most common collusive content is that the industry has reduced supply or capacity and refers to competitors or the industry to facilitate such conduct. This type of content is present in 37% of the 109 true positive transcripts. In emphasizing supply or capacity cuts, a firm may highlight how those reductions are in the industry's best interests. For example:[21]

> "We've seen a massive ... consolidation in the industry ... and hopefully, that leads to a better, a healthier competitive environment as well. No one is going to build DRAM capacity for the foreseeable future. It's just not going to happen."
> [Micron Technology, Inc. | Company Conference Presentations | 05 Dec 2012]

> "Other companies in our industry have announced considerable domestic downtime and capacity closures [and] share my philosophy that if you can't sell it or you can't use it, you shouldn't make it."
> [WestRock Paper and Packaging, LLC | Earnings Calls | 11 Nov 2008]

> "We will not continue to invest at this pace, and our activity will slow. It's kind of a no-brainer from an economic standpoint. Our partners are making the right decisions in terms of the decreasing capital in this environment."
> [Ultra Petroleum Corp. | Earnings Calls | 03 May 2012]

[21] For each quoted excerpt, we include in square brackets the company name, event type, and date, as reported in Compustat, corresponding to the original transcript that was provided in the LLM query. While previously documented cases like airlines are in the set of true positives, we chose not to use excerpts from them (even though some are highly illustrative) in order to focus on new episodes identified by the LLM. As the passages are from the LLM-generated excerpts, the sentences may be from different parts of the transcript and not appear together as they are presented here.

In some cases, the firm went beyond noting a decline in supply or capacity to express more reductions are needed.

> "We currently anticipate these conditions will require the industry to rationalize production to stabilize the markets so fundamentals can improve. While some supply response has already occurred, more reductions are necessary to balance the market."
> [Alliance Resource Partners, L.P. | Earnings Calls | 28 Oct 2019]

> "On the market side, [there is] still [a] challenging environment in Europe. Industry-wide capacity closures are supportive [and] if we can see maybe 1 or 2 more machine closures in Europe, I think then we are more or less moving it."
> [Norske Skogindustrier ASA | Earnings Calls | 07 Feb 2013]

> "There continues to be too much capacity in the global preclinical services industry, and the excess capacity keeps pressure on pricing. We recognized that we would have to reduce capacity in order to improve utilization. A competitor has just announced capacity reductions, and we expect that the pharma industry will continue its efforts to sell and/or close preclinical space. We believe that more capacity remains to be rationalized [and] these actions will help to improve the pricing dynamics."
> [Charles River Laboratories International, Inc. | Earnings Calls | 04 Nov 2010]

In nine of the 40 transcripts, reduced industry supply or capacity is mentioned along with references to competitors doing so and the firm supporting such conduct.

> "We are limiting our investments with reductions in capital [and] production declines will soon be evident. We are witnessing peer companies follow our lead [and] think more companies should adopt this prudent practice."
> [Ultra Petroleum Corp. | Earnings Calls | 02 Aug 2012]

> "Industry linerboard and Kraft paper capacity will be substantially reduced [and] we should expect to see further capacity closures announced in the coming months. [This] should lead to substantial price restoration [as] high operating rates give the industry courage to raise prices. I believe that's good for our shareholders and good for the industry."
> [WestRock Paper and Packaging, LLC | Earnings Calls | 05 Nov 2009]

> "The industry is oversupplied in NAND, and we need to slow down bit supply. So everybody is slowing. I think the industry is reacting well. We're not trying to gain share. It's about margin profile and profitability."
> [Western Digital Corporation | Company Conference Presentations | 29 Nov 2022]

And when competitors have not reduced supply or capacity, the firm calls them out for not doing so.

> "We've also done what we think is the prudent thing to do [which] is to effectively provide ... incremental storage to the market by curtailing our production and not pulling it out of the ground. It has been a mystery to me why more producers have not made that same economic choice. We think that it has a chance to restore gas prices, but we'll see how that pans out."
> [Expand Energy Corporation | Company Conference Presentations | 27 Mar 2012]

In some instances, a firm notes that it is acting as a market leader in reducing supply or capacity. With that leadership reference comes the implicit invitation for competitors to be followers.

> "We are currently idling temporarily ... approximately 40% [of] operating capacity [and] we will be implementing a price increase of up to 10%. There is a supply and demand imbalance in this business, and being a market leader, we have taken a step to reduce our capacity."
> [Ampco-Pittsburgh Corporation | Earnings Calls | 27 Feb 2015]

> "We took our production level down [and] we believe our actions had a positive effect on bringing industry ethanol stocks into a more stable supply and demand balance. Somebody had to exert leadership in this industry to do it. And if we did not slow down, we would be in a much worse margin situation today."
> [Green Plains Inc. | Earnings Calls | 01 Aug 2017]

> "We have made production curtailments. Following our lead, we're now seeing a number of recent capacity curtailments by other producers as the industry takes action necessary to support a recovery in prices. We view this trend favorably and see this dynamic as critical to stabilizing and eventually improving the pricing environment."
> [Ferroglobe PLC | Earnings Calls | 03 Dec 2019]

5.3 Category 2: The industry needs to reduce supply or capacity or to show discipline

The second most common category is when a firm announces that the industry needs to reduce supply or capacity or to show discipline. These instances made up 28% of the true positives. As documented in airlines and steel (see Aryal et al. (2022) and Harrington (2022)), "discipline" is a common code word for "less competition". Other loaded words include firms being "responsible" and "rationalizing" capacity.

"We do consider the possibility of reducing our production in 2017 in order to have a better market environment and also a better environment for pricing. If we in the industry … do not have a discipline regarding our supply, ... we could be generating a situation in which the industry's ROIC could be lower. We are not sure that this will bring other companies to follow in terms of the supply strategies. But what is clear to us is that [we have] the responsibility of showing the facts to be followed in order to seek profitability for the whole industry."
[Suzano S.A. | Earnings Calls | 26 Oct 2016]

"It is important that the U.S. oil and gas industry as well as other global producers continue exerting capital discipline to not overproduce. Everyone needs to continue to cutback and participate as this slump continues. We reduced our pace of growth to deliver a disciplined approach to value creation."
[Continental Resources, Inc. | Earnings Calls | 11 May 2020]

"We plan to optimize near-term pricing through the voluntary curtailment of approximately 100 million cubic feet per day over the next several months. Our hope is that our peers, some by choice rather than necessity, will do the same. In today's market, I believe what our industry needs is a more measured pace of growth."
[Gulfport Energy Corporation | Earnings Calls | 05 Nov 2015]

One firm went so far as to specify how much capacity needed to be cut.

"It's an obvious conclusion that scrapping has to take place [and] according to my calculations, about 1/3 of the global fleet should be scrapped. Consolidation is inevitable and will take place in our industry."
[Prosafe SE | Earnings Calls | 27 May 2020]

These expressions of an industry need to cut supply or capacity may be explained in the context of raising prices.

"The average prices over the last 15 years in Europe has been EUR 550 … and we believe that ... the trend prices should be more in this range than the historically very low levels that we've seen this year of EUR 420, EUR 430."
[Norske Skogindustrier ASA | Earnings Calls | 04 Nov 2010]

The firm may go on to note being encouraged by competitors' conduct.

"The market has overcapacity now, which puts the leverage in the customers' hands. You can resolve overcapacity by not making investments. What needs to happen is capacity needs to shrink. I've been encouraged by what I see our competitors doing."
[Superior Energy Services, Inc. | Company Conference Presentations | 12 Feb 2019]

Especially egregious is when a firm announces how its own supply or capacity cuts reflect "doing its part by exerting discipline" [352688] and then demands competitors to pull their weight.

> "We've implemented planned supply discipline. We think we've done our share and maybe more."
> [Cameco Corporation | Earnings Calls | 04 Nov 2020]

> "And candidly, we can't do it alone. It's got to be an industry level. There's a gross oversupply in the space. We're taking aggressive action on the supply side."
> [Micron Technology, Inc. | Company Conference Presentations | 29 Nov 2022]

> "The market is clearly oversupplied. We are reducing capital by nearly 20% and production approximately 15%. One company is not positioned to 'fix the market'."
> [Expand Energy Corporation | Earnings Calls | 21 Feb 2024]

> "We proactively impaired and early retired a number of our older coastal barges and tugboats. We believe this action is necessary not only to help restore an early balance into the market, but also to improve our profitability going forward. If others will follow our lead, an accelerated balance can be restored to the struggling market. We did our part to rationalize the industry's fleet and improve our profitability going forward, but we need the industry to mirror our actions."
> [Kirby Corporation | Earnings Calls | 01 Feb 2018]

5.4 Category 3: The industry has shown or will show supply or capacity or price discipline

Rather than expressing a need for more supply or capacity discipline, a firm may note that it has occurred or will occur and expresses support for it. This or similar language appeared in 11% of the true positive transcripts. These transcripts note the firm's supply or capacity cuts in the context of an industry effort.

> "Market conditions remain challenging, exacerbated by the overall oversupply in the industry. We started maintenance of our facilities and adjusted our production utilization rate to 50% in light of weak market demand. There has been consensus among the manufacturers to promote a healthier development of the industry through exercising self-discipline."
> [Daqo New Energy Corp. | Earnings Calls | 30 Oct 2024]

"We have seen proactive initiatives to restore the industry's healthy development. Implementing sales discipline would be fundamental to mitigating the irrational competition amid falling prices. We plan to maintain a relatively low utilization rate in 2025 until a turning point emerges in the sector."
[Daqo New Energy Corp. | Earnings Calls | 27 Feb 2025]

"If we could scare people out, and they knew how much we were going to be building [then] maybe they wouldn't build theirs. We were trying to get some attention and show that, hey, everybody else should be disciplined, and we're going to come out and throttle this. This is not a cycle where you would expect to see a tremendous amount of capacity come on."
[ProFrac Holding Corp. | Company Conference Presentations | 07 Sep 2022]

A firm may commend competitors for acting responsibly with respect to their output or prices.

"We finally have sanity back in the seaborne iron ore market. I truly commend Rio Tinto and Vale for eliminating the reckless behavior that had infected the market for a number of years. We expect this type of responsible behavior to carry on, which will continue to support strong iron ore prices."
[Cleveland-Cliffs Inc. | Earnings Calls | 09 Feb 2017]

"The entire industry moved to action and acted responsibly last year. So we were able to, for the first time, put the prices up and that's a great thing for our business and for our industry."
[Coca-Cola Bottlers Japan Holdings Inc. | Earnings Calls | 10 Feb 2023]

Based on a recent change in market structure, a firm may forecast future discipline.

"We feel that consolidation somehow will improve market discipline going forward. Given the fact that operating in such a competitive scene has led to price erosion in data service pricing, improving market discipline is a very welcome element and consequence of this transaction."
[PT XLSMART Telecom Sejahtera Tbk | Special Calls | 26 Sep 2013]

5.5 Category 4: The industry needs to raise prices

Rather than communicating that the industry needs lower supply or capacity, a firm may announce that it needs higher prices. Such was present in 11% of the true positive transcripts.

"Tariff repair is sorely needed for return ratios to improve. We could do it, we could lead it, we could start it, but the fact is that if competition doesn't follow, then it will hurt us. All the players need it."
[Bharti Airtel Limited | Earnings Calls | 15 May 2024]

> "I think it's not only our desire to generate that level of profitability. So the need of increasing pricing, I think, will be more an industry need. We think we've arrived at a time where it was necessary to increase a little bit our pricing."
> [Seagate Technology Holdings plc | Company Conference Presentations | 05 Sep 2023]

> "We have an intention to do further price increases. The entire industry has the same need. Otherwise it wouldn't work. So there is a strong need for price increases in Consumer Tissue in Europe. Many of our competitors in Europe have lower margins than we have to start with. So they are in as much need of price increase as we are or even more so."
> [Essity AB (publ) | Earnings Calls | 19 Jul 2018]

The called-for price increases may be in the context of responding to cost increases or ending a price war.

> "We as an industry need to do a better job of tying those contracts to a CPI that actually means something to us, right? We can't be the volume leader by stealing from our competition because we know what that's going to lead to. But in a market that's growing, we should get our fair share of volumes."
> [Waste Management, Inc. | Company Conference Presentations | 13 Nov 2014]

> "Price war destroys value, and it should be reversed soon. If you look at the structural positioning of Telecom Italia, we are by far the strongest player. And therefore I think that the others are suffering as much and more than we are suffering. So it is not something that can be sustainable. This price war is unsustainable for all the players. I consider it a missing opportunity for the whole Italian market."
> [Telecom Italia S.p.A. | Earnings Calls | 02 Aug 2013]

5.6 Category 5: The firm is being or will be less aggressive in pricing and references competitors' conduct in a manner encouraging similar conduct

In eight transcripts, the firm announced it has raised price or will raise price and, when referencing competitors' pricing, encouraged them to do so or commended them for having done so.

> "Our strategy of increasing the prices during quarter one has paid off. We [are] going to increase prices further in second half [as] we believe that there is still room to go in that direction. We hope that the market goes in the same direction. The market is now rational, and we believe that there is room for it to keep improving."
> [PT XLSMART Telecom Sejahtera Tbk | Earnings Calls | 28 Jul 2023]

"We don't want to adjust prices down. We are not cutting prices further. We have been able to maintain our prices. We have seen signs that some of our competitors are doing the same."
[Embraer S.A. | Company Conference Presentations | 08 Jun 2016]

"Our primary objective will be optimizing profits. We made a conscious decision to trade off attendance [i.e., volume] for yield. We can lean even more heavily into pricing and expect this will further lower our attendance. I can congratulate SeaWorld and Cedar Fair for how they have maintained better pricing integrity than us."
[Six Flags Entertainment Corporation | Earnings Calls | 12 May 2022]

The firm may recognize its role as a market leader with regards to pricing.

"We adopted a very clear strategy of price-over-volume strategy for the bromine [and] we are acting as a market leader. When there is a need, we are reducing ... our volumes in order to keep the prices at the right level. We don't want prices of bromine as the leader in the market to be too high because this will start to create incentives for other producers. We believe that the current price level of the bromine is more or less optimal."
[ICL Group Ltd | Company Conference Presentations | 30 Nov 2016]

5.7 Some Other Examples

Some other examples include a firm putting forth an industry objective that puts more weight on price and margins than on market share and volume. If adopted by firms, it has the implication of less aggressive competition and higher prices.

"Not only Seagate but the entire industry, decide to focus more on improving the bottom line than fighting for market share. Market share is not our focus. It's taking the right profit from this product. Industry could be constrained. Our objective is to be very reasonable with the capacity we put in place."
[Seagate Technology Holdings plc | Company Conference Presentations | 29 May 2024]

"Value over volume. The industry is not really equipped to really capture that value proposition. We're going to have to start, as an industry ... to collaborate on turning this dynamic around [by] consolidating districts, rationalizing capital expenditures, not building duplicate infrastructure in the same camps."
[Goldcorp Incorporated | Company Conference Presentations | 16 May 2017]

Finally, one transcript facilitated hub-and-spoke collusion as a manufacturer-hub sought to dampen competition among dealer-spokes.

"We will not oversupply the market [and] we hope that our dealers will follow suit and stop competing with themselves."
[Harley-Davidson, Inc. | Earnings Calls | 28 Jul 2020]

6 Some Implications for Enforcement

The initial evidence presented here supports a valuable role for an LLM in a screening process. An LLM is effective in pulling out those rare public announcements with collusive content from a massive set of company transcripts. It is able to effectively substitute for human experts in performing an initial screen. However, given the high rate of false positives and the importance of a competition agency avoiding the allocation of resources to false leads, human oversight of the LLM's output is critical. Here is a candidate protocol:

Step 1: LLM reads and rates a corpus of transcripts.

Step 2: For those transcripts with the highest scores, human experts read the LLM-generated excerpts and identify those with collusive content.

Step 3: For a transcript surviving step 2, human experts read the entire transcript and determine if it is worthy of further examination.

Step 4: For those deemed worthy of further examination, all transcripts are collected for the company along with competitors' transcripts and read by human experts.

Based on the assessment of that body of transcripts, it can be determined if an official investigation should be initiated. The case study in the next section provides proof of concept as an LLM-based screen with a very modest use of human resources is able to identify a compelling case for prosecution.

Prior to pursuing cases based on public communications, it would be prudent for a competition agency to release guidelines describing the content that companies should avoid in their public communications. While courts have recognized that public communications can result in an unlawful agreement, exclusive reliance on public communications to obtain a conviction is relatively unexplored territory. Issuing guidelines lays the legal groundwork for public prosecution or private litigation as well as deterring collusive conduct. As described in Harrington (2022), a firm's public announcement risks facilitating collusion when it refers to the conduct of rival firms or the industry large. Our categorization of the content of the LLM-identified transcripts provides the basis for guidelines describing the various types of content that firms should avoid in their public statements.

7 Case Study

The case study examines the domestic airline industry in Brazil from 2012 to 2016. GOL Linhas Aéreas Inteligentes (GOL) and LATAM Airlines (LATAM) were the dominant airlines throughout this period.[22] The other two key airlines are Aerovías del Continente

[22]Effective June 2012, Brazilian Airline TAM Linhas Aéreas purchased Chilean LAN Airlines to form the LATAM Airlines Group. Thus, there will sometimes be reference to LAN Airlines and TAM.

Americano (Avianca) and Azul Linhas Aéreas Brasileiras (Azul) who, while starting small in 2011, saw their market shares grow significantly. Figure 10 reports the four airlines' capacities over 2011-16.

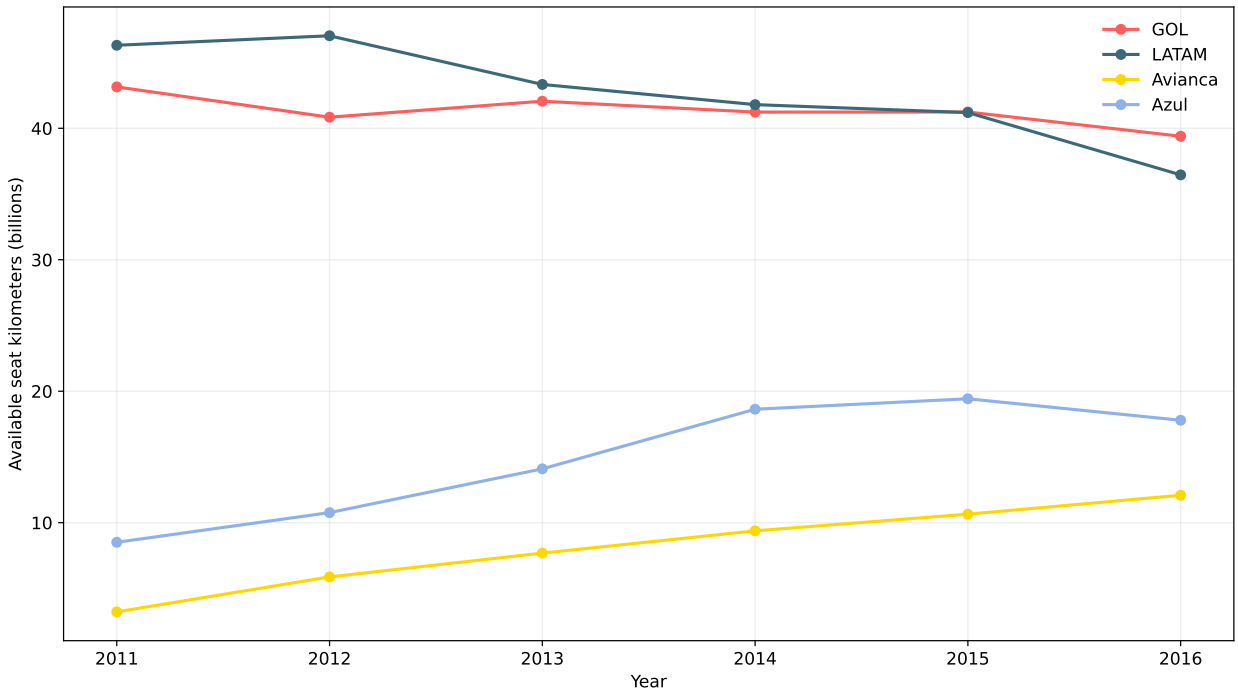

Figure 10: Airline capacity in Brazil, 2011–16

Notes: Available seat kilometers (ASK) for GOL, LATAM, Avianca, and Azul on domestic regular flights, aggregated to airline-year totals from ANAC data Agência Nacional de Aviacao Civil (ANAC) 2011–2016.

The time period 2012–16 was selected based on the LLM's output. Over the entire database spanning from 2002 to 2025, there are 25 LLM-flagged GOL transcripts (score of at least 75) and 20 of those occurred in 2012–16 with only one over 2017–21 and four during 2022–23.[23] The LLM-validated step screened 15 GOL transcripts (with a mean score of at least 75 based on at least 10 runs) and all but one of them occurred during 2012–16.[24] For LATAM, it had eight LLM-flagged transcripts with six of them over 2012–16 (of which three are LLM-validated) and two in 2017 (of which one is LLM-validated). In sum, the LLM is identifying a large number of suspicious announcements during 2012–16, none in the preceding years, and very few in the immediately ensuing years.

We limited our analysis to the 39 earnings calls for GOL and LATAM over 2012 to

[23]LLM-flagged transcripts for GOL: 2012 (4 transcripts), 2013 (7), 2014 (4), 2015 (2), 2016 (3), 2017 (0), 2018 (1), 2019 (0), 2020 (0), 2021 (0), 2022 (3), 2023 (1).

[24]LLM-validated transcripts for GOL: 2012 (4 transcripts), 2013 (5), 2014 (1), 2015 (2), 2016 (2), 2022 (1).

2016.[25] During that time period, there were no earnings calls for Avianca and Azul.[26] Provided below in chronological order are all of the earnings calls with any relevant content. When the content is conveying a similar message, they are placed together under a bullet point with a one or two sentence description of that message. For each earnings call, a succinct description of the content of the earnings call is provided along with the date (in the format year/month/day). For example, GOL earnings calls for March 27, 2012 and May 4, 2012 both convey that GOL is limiting its supply. So that the reader can draw their own interpretations, relevant excerpts from the LLM-identified passages are provided in the footnotes.[27]

- GOL states it will limit supply.
  - 2012/03/27 – GOL: We are committed to responsibly adding supply. [28]
  - 2012/05/04 – GOL: We are reducing supply and the industry will restore profitability by limiting supply.[29]
- GOL and LATAM express the industry needs to cut capacity and note that they are cutting capacity but some firms are not cutting capacity.
  - 2012/08/13 – LATAM: The market has suffered from overcapacity but conditions are in place for the industry to limit capacity. We will decrease capacity.[30]
  - 2012/08/14 – GOL: We are increasing our capacity reduction and we expect competitors to be less aggressive in pricing. But our market share has been declining because some firms are expanding.[31]

[25] For purposes of brevity, we are excluding transcripts that are not earnings calls. Research assistant Vivian Zhang read the entirety of the 39 earnings calls and collected excerpts with possible collusive content. Co-author Joseph Harrington then read the excerpts upon which the analysis of this section is based.

[26] Due to when Avianca was listed on the NYSE and various restructuring, the database has 11 earnings calls from 2017 Q2 to 2019 Q4 and nine earnings calls from 2023 Q1 to 2025 Q1. Azul was not listed on the NYSE until 2017 and the database has 33 earnings calls from 2017 Q1 to 2025 Q1.

[27] To ease reading, many excerpts have been condensed and combined but always in a manner not to affect their meaning. In cases where the meaning of the content is unclear, the original passage was reproduced unaltered.

[28] "The company has already affirmed its commitment to the responsible addition of supply on the domestic market and recently announced a 0% supply growth. We are therefore announcing a capacity cut as of March."

[29] "GOL's new target of reducing domestic supply by up to 2% over 2012, together with the industry adoption of a conservative approach to increase supply, are indicative of a market governed by rationality in order to resume profitability. The long-term winners are the ones that have the courage of not saying that they are buying a lot of planes in the short run."

[30] "Brazil has historically suffered from overcapacity. However, we believe that conditions are in place for carriers to sustain capacity ceiling and focus increasingly on profitability. We currently expect decreased capacity between 2% and 3% this year."

[31] "Previously, we're saying that the decrease in capacity would be 2% for the year. And now, we are saying that the new range is between 2% and 4.5% on the negative side. We do expect to see our competitors

  - 2012/11/13 – LATAM: Some firms are pursuing capacity discipline and we support it for the industry.[32]
  - 2012/11/14 – GOL: We and a competitor have cut capacity and are showing capacity discipline, and we intend to cut more.[33]

- GOL and LATAM emphasize that capacity discipline will restore profitability, and GOL notes that other firms are now following its lead to reduce capacity.

  - 2013/03/20 – LATAM: We believe capacity discipline is the key to restoring profitability.[34]
  - 2013/03/26 – GOL: Other firms are now following our lead to reduce capacity. Less capacity is the path to higher margins.[35]

- GOL and LATAM reiterate that reducing capacity is the key to better market outcomes—such as higher revenue—and they are continuing to practice capacity discipline.

  - 2013/05/15 – LATAM: We believe capacity discipline is the key to higher profits.[36]
  - 2013/08/21 – LATAM: We continue to focus on reducing capacity. [37]
  - 2013/11/13 – GOL: Supply is down and that is key to improving market outcomes.[38]

---

being less aggressive in their price policies for the following months. But when we see the total market, it's also clear that our market share in combination with our main competitor, is going down because the newcomers are investing a lot in increasing their capacities."

[32] "Some of the major players of the industry are taking a strategy of capacity discipline that will help to increase the load factor. We're going to listen to capacity discipline of the industry."

[33] "We cut the capacity on April in comparison to March by 8%, which is going to give us an annual effect of around 4.5%. The competitor just now is taking the same measure. We do believe that the current industry discipline should reduce capacity and we would like to find additional opportunities to reduce capacity."

[34] "We remain convinced that capacity discipline and adequate segmentation of the market will provide the basis for continued healthy load factors and significant improvement in operating results in 2013. Our strategy is very focused on capacity discipline, and we believe that this is key for regaining profitability for the year in Brazil."

[35] "For 2013, the company has already announced the curbs of domestic supply of around 7%, meaning a reduction of more than 10% in 2 years, reinforcing the company's commitment to resuming operating margins. The leading market players have begun to follow the supply reduction initiated by GOL in April 2012, an important sign that Brazil's airline industry is on the way to constructing a more sustainable environment."

[36] "We remain convinced that capacity discipline and adequate segmentation of the market will provide the basis for continued healthy load factors and expect continued improvement in operating results in 2013."

[37] "We continue focused on reducing capacity. We expect to reduce capacity by between 7% and 9%."

[38] "Domestic supply is down year-over-year by 7%, almost 10% in the 9 months of 2013. We think this is absolutely necessary for the market to reorganize itself."

  - 2014/03/18 – LATAM: Maintaining capacity discipline has led to higher profits.[39]
  - 2014/03/26 – GOL: The priority is raising PRASK (passenger revenue per available seat kilometer) and a main driver is reducing capacity and rationalizing supply. We do not see future industry capacity additions. [40]

- GOL and LATAM say they are staying the course of capacity discipline, and LATAM notes that others are expected to do so, too.
  - 2014/05/14 – LATAM: The competitive environment remains intense and we continue to support a rational capacity strategy.[41]
  - 2014/05/15 – GOL: We are rationalizing supply to pursue price increases.[42]
  - 2014/08/13 – LATAM: We reduced capacity and expect others to do so, too.[43]
  - 2014/08/14 – GOL: The industry is pursuing limitations on capacity. [44]
  - 2014/11/12 – GOL: We will continue to be rational in terms of capacity.[45]

- GOL suggests that Avianca and Azul will engage in capacity discipline.
  - 2015/03/31 – GOL: We believe Avianca and Azul will pursue the rational sales approach started by GOL and followed by LATAM.[46]

[39] "TAM continues to make significant improvement in the financial result of the domestic Brazil passenger operations, maintaining capacity discipline, 6% reduction quarter-over-quarter."

[40] "Our strategy of rationalizing supply is in place and is one of the main drivers that is causing our revenues to really move up. The priority is to increase the PRASK at the level of 2 digits. So that strategy has been quite successful, and it has also helped the markets to drive towards better capacity discipline. We do not see in the horizon, the time when the Brazilian domestic market will be able to accept additional capacity again."

[41] "We do expect to see continued reductions in ASKs during the second quarter. The competitive environment in Brazil remains very intense. So we continue with very rational capacity strategy in the domestic Brazil with a very positive effect on load factors."

[42] "We can see GOL's strategy to pursue an increase in price and rationalization of supply which is currently 12% lower than when the company started to cut capacity in May 2012."

[43] "What we observed is that there's a lot more capacity additions than demand. We reduced this capacity, and we expect others to reduce."

[44] "The industry is still moving towards a very strict control on capacity. The indications we got from the market related to the competitors and to the industry as a whole doesn't show any important movement to add capacity. Therefore, I think the (pricing) discipline will continuously offer us a healthier environment to recover margins."

[45] "We will still be very, very rational in terms of capacity in Brazil. The strategy has been working for us."

[46] "We are maintaining our firm commitment to not expanding domestic supply in 2015. But I could say that we do see a more rational approach in the future sales eventually coming from the market as a whole. We do expect that our competitors will follow this rational approach already started by GOL, followed by TAM. And we believe that it should be followed by Avianca and Azul along the following months. We say our roof (for capacity) is 0 growth. It could be lower. This is basically our clear signage to the market that we won't increase capacity."

  - 2015/05/13 – GOL: We continue to be committed not to expand supply. While some competitors have increased supply, we believe this is about to change.[47]

- GOL states that some firms have not followed their lead to reduce capacity. Both GOL and LATAM say they will continue with capacity cuts.

  - 2015/08/14 – GOL: We led in cutting capacity but some competitors did not and there is still excess capacity. We will continue to cut.[48]
  - 2015/08/14 – LATAM: We are cutting capacity more than originally planned.[49]

- GOL expresses that other firms seem to be ready to reduce capacity, and LATAM notes that more capacity needs to be cut.

  - 2015/11/12 – GOL: The industry is now showing signs of support for capacity reductions.[50]
  - 2015/11/13 – LATAM: We were the only firm to reduce capacity in the last quarter. More capacity needs to be cut.[51]

- GOL and LATAM continue to support capacity reductions.

  - 2016/03/09 – LATAM: We have been leading the industry in capacity cuts and capacity reduction is occurring throughout the industry.[52]

[47] "We maintain our commitment of not expanding the supply in the domestic market. There is an over-capacity because some of our competitors have increased their capacity over the last period, while the demand came down. I believe that this behavior is about to change when we look at the future sales – the inventory. I would say all the competitors have changed the number of seats being pushed available for sales along the following months. And this is a quite good signal that this capacity discipline will be considered more seriously by everyone else."

[48] "We still do have an over-capacity in the market. Therefore, we are again leading the market to further capacity decrease. We will reduce supply to between 2% and 4% in the second half. We have started this capacity reduction already in April 2012. And we have cut, since then, 14%. LATAM is somehow catching up the same capacity reduction."

[49] "We were previously expecting to be flat for 2015 in terms of total capacity and our current estimate is to have a reduction of between 2% and 4%."

[50] "The industry is giving clear signals that there are signs that there is a different behavior related to capacity. The major difference in comparison to the first capacity reduction cycle is that now the pressure on the demand does not allow any capacity increase even from the other competitors."

[51] "TAM was the only company to reduce capacity in the third quarter of this year. We see a complicated economic year for Brazil. That's why we're forecasting another reduction for next year as necessary. Our estimate is to reduce capacity between 9% and 6% for next year."

[52] "And we have been leading the industry in terms of capacity reduction. We had originally talked about a reduction of between 6% and 9%. Now we have increased that to a reduction of between 8% and 10%. What we do see is the capacity reduction overall in the industry, which we think will have positive impacts in our yields in the future."

- 2016/03/30 – GOL: Our supply cuts have been rational for the industry.[53]

- GOL and LATAM state that they have cut capacity but more needs to be cut. GOL notes some firms have still not cut capacity.

  - 2016/05/12 – GOL: In response to overcapacity, we have reduced capacity but small firms have not. There needs to be more capacity cuts. [54]
  - 2016/05/12 – LATAM: We are increasing our capacity reductions but more industry capacity needs to be cut.[55]

- GOL expresses how it has led in reducing capacity, some firms have not followed, and all firms benefit from reduced industry capacity. LATAM conveys that there is too much industry capacity and it will not add capacity.

  - 2016/08/12 – LATAM: We do not expect to raise capacity.[56]
  - 2016/08/16 – GOL: We have led in reducing capacity but two firms have not followed that strategy. We will continue to cut capacity and believe lower industry capacity will benefit all firms.[57]
  - 2016/11/07 – GOL: We have led in reducing capacity but two firms have not followed that strategy. We will continue to cut capacity and believe lower industry capacity will benefit all firms.[58]

[53] "Together, we are responsible for approximately 14 billion ASKs less in the system that has been bringing more rationality for the industry."

[54] "Overcapacity in Brazil has added pressure on GOL. While the company has been very disciplined, others have not. GOL has been reducing capacity since 2012. However, smaller players have been adding capacity, negatively affecting sector results. The discipline regarding capacity is much more rational than before, but I wouldn't say that it has achieved its balance in comparison to the demand when we are talking on the whole system. It's improving but we are still not close to the optimum point."

[55] "We are increasing our capacity and we now expect capacity to be reduced between 10% and 12%. We're convinced that this rational approach to capacity is key to being able to improve our results in that market. Despite the strong capacity cuts that the market implemented, it was still not enough to compensate for the demand drop."

[56] "We continue to expect our passenger capacity to be relatively flat this year with respect to 2016."

[57] "Although GOL has provided leadership in reducing capacity, industry profitability has been negatively impacted by the capacity growth of 2 players, one of which has not reduced capacity, and the other has even added capacity during the downturn in the market. The competitive environment is still tough. We also believe that the reduction in the industry's capacity will benefit us all."

[58] "Since 2011, GOL has worked to lead the industry in rational and profitable capacity growth. In the adjustment of industry capacity to match demand, not all competitors have cut capacity and reduced frequencies on destinations and routes in Brazil. Some competitors have continued to cut capacity and reduced frequencies. Other competitors have continued to add capacity, in spite of their higher operating costs. And this has had a big impact on sector profitability. To the extent that everybody is disciplined on capacity, we should see a solid yield environment."

– 2016/11/11 – LATAM: There is still too much capacity and we will not add capacity in the near term.[59]

During public announcements spanning from 2012 to 2016, GOL and LATAM explicitly and repeatedly stated there is too much industry capacity and reassured each other that they will reduce capacity. Good to their word, both airlines significantly reduced their capacities as can be seen in Figure 10. GOL provided confirmation in their 2016/03/30 earnings call.

> "Along the last 4 years ... GOL reduced approximately 14% of its supply in the domestic markets and was followed by its main competitor. Together, we are responsible for approximately 14 billion ASKs less in the system, a phenomena that has been bringing more rationality for the industry. However, it's not the same effect produced by other competitors in our industry ..."

As just noted by GOL, Avianca and Azul did not reduce their capacities and GOL and LATAM criticized those two airlines for not doing their part. Perhaps to encourage their cooperation, GOL conveyed the expectation that Avianca and Azul were going to get on board, though apparently they never did. The explicitness of the content and the repeated back-and-forth between GOL and LATAM is compelling evidence that they both invited coordinated capacity cuts and both accepted the other firm's invitation. In other words, they had an agreement to reduce their capacities which, based on this evidence, is likely to be an antitrust law violation in many jurisdictions.[60]

The case study illustrates the potential value of screening public announcements using an LLM. The LLM's screen identified public announcements for GOL and LATAM containing collusive content. This finding was the basis for conducting a human review of public announcements by GOL and LATAM during the time period identified by the LLM. That human review delivered compelling evidence of communications facilitating an illegal agreement. If it had been found at the time, we believe this evidence would have justified prosecution of GOL and LATAM by the Brazilian competition agency Conselho Administrativo de Defesa Econômica (CADE). [61]

[59] "We still believe that there is some excess capacity in the market. We're not thinking about adding any capacity in the next few months."

[60] Whether the capacities of GOL and LATAM were lower because of these announcements is a separate question about economic effect rather than liability. To say something about the counterfactual capacities, one would need to engage in an analysis such as in Aryal et al. (2022). They used variation in the content of earnings calls across markets and time to test whether all (legacy) airlines on a route having "capacity discipline" or similar content in their earnings calls in a quarter was correlated with lower capacity for that route in the ensuing quarter.

[61] Inspection of earnings calls from 2022–23 shows a resurgence of disconcerting content. There were four LLM-flagged GOL transcripts and three of them had relevant content. "GOL have led the market towards rationality. The industry has demonstrated an even higher level of rationality in comparison to what we have seen so far." (2022/03/14, mean score = 74.17) "We're being very disciplined on capacity given what's going on with costs and particularly fuel costs. I can say it was an industry cut because we also put 10%.

8 Concluding Remarks

This study contributes to addressing two questions: How do firms collude? And how do we detect collusion? The focus is on how firms coordinate their conduct with an underexplored communication practice: public announcements such as earnings calls. To address how public communications could facilitate collusion, an LLM is deployed to find company transcripts with collusive content. The LLM identifies many new episodes which vastly adds to the few episodes discovered through public and private enforcement (Harrington 2022). A human audit of those episodes reveals some regularities in the messages used to coordinate an output or capacity reduction or a price increase. Consequently, we are left with a more informed understanding of how public announcements can be a cause for anticompetitive concern. The human audit is also the basis for addressing the second question as it is assesses the performance of the LLM as a screen for collusion. The LLM is able to substitute for human experts in finding those rare transcripts with collusive content though the rate of false positives means there is an essential role for human oversight.

We view our study as an initial step of a research program investigating public communications and collusion with the assistance of generative AI. Moving forward, let us close with some thoughts on improving the performance and how the output of our study can be used for new research projects.

There is much more that can be done with an LLM. We have had it evaluate each company's transcript in isolation. Having an LLM simultaneously evaluate a collection of transcripts could enhance performance. This could be done by considering all competitors' transcripts and have it attach a higher score when collusive content is more common among competitors. That is a feature of the approach taken in Duso et al. (2025) with a dictionary model. Alternatively, an LLM could consider the entire collection of a company's transcripts and have it look for changes in the extent of collusive content. While that could be performed with the existing model - by looking at scores over time - it may be more effective in having an LLM look for the arrival and departure of collusive content. A second direction pertains to evaluating performance. The human audit examined the transcripts with the highest LLM scores and thus those identified as suspicious. While the audit then informed us of the rate of false positives, it could not speak to the rate of false negatives; how many episodes is the LLM failing to pick up? To do so would mean extending the human audit to some of the cases with non-high scores. That may require many more human resources but could provide valuable information about an LLM's performance.

Our study has focused on identifying and synthesizing the content of public communications that facilitate collusion. A next research step is to explore whether it results in

We saw also competition cutting 10% to sustain all the fare activity that we are required to do in this very high cost environment. What we are seeing in the market is really a rational behavior in terms of fares." (2022/07/28, mean score = 76.25) "We want to create the right equilibrium and demand and supply. We are positioning ourselves as kind of a leading the capacity discipline, we are cutting. We are seeing that there is rationality in the market." (2023/07/27, mean score = 65)

effect along the lines of the analyses in Aryal et al. (2022) and Sheng and Vukina (2024). Our analysis has discovered many candidate markets for conducting such an empirical analysis. Two necessary elements of an episode is that it has a sufficiently rich set of public announcements - in particular, variation in the presence of collusive content - and data is available on outputs, capacities, or prices so as to measure effect.

In concluding, we believe an LLM has proven it is street-ready to be used by competition agencies and plaintiff law firms to search for episodes warranting investigation and possible prosecution or litigation. For that purpose, we have proposed a protocol combining the LLM with human oversight. It is our hope that this study and that of Duso et al. (2025) are just the beginning of the development of an effective enforcement tool.

References

Aaltio, Aapo, Riku Bari, Antoo Jokelainen, and John Lundberg (2025). "Complementary Bidding and Cartel Detection: Evidence from Nordic Asphalt Markets". *International Journal of Industrial Organization* 98, 103129.

Agência Nacional de Aviacao Civil (ANAC) (2011–2016). *Air Transport Market Statistics: Aggregated Supply and Demand Data (Demanda e Oferta – Dados Agregados)*. Brasília: ANAC. URL: https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/dados-estatisticos/dados-estatisticos.

Aryal, Gaurab, Federico Ciliberto, and Benjamin T. Leyden (2022). "Coordinated Capacity Reductions and Public Communication in the Airline Industry". *The Review of Economic Studies* 89, 3055–3084.

Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer (2026). *GPT as a Measurement Tool*. NBER Working Paper 34834. National Bureau of Economic Research. DOI: 10.3386/w34834. URL: https://doi.org/10.3386/w34834.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016). "Measuring Economic Policy Uncertainty". *The Quarterly Journal of Economics* 131.4, 1593–1636. DOI: 10.1093/qje/qjw024. URL: https://doi.org/10.1093/qje/qjw024.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3, 993–1022. URL: https://www.jmlr.org/papers/v3/blei03a.html.

Borenstein, Severin (2004). "Rapid Price Communication and Coordination: The Airline Tariff Publishing Case (1994)". *The Antitrust Revolution*. John E. Kwoka and Lawrence J. White, eds. Oxford University Press.

Bourveau, Thomas, Guoman She, and Alminas Žaldokas (2020). "Corporate Disclosure as a Tacit Coordination Mechanism: Evidence from Cartel Enforcement Regulations". *Journal of Accounting Research* 58, 295–332.

Byrne, David P. and Nicholas de Roos (2019). "Learning to Coordinate: A Study in Retail Gasoline". *American Economic Review* 109, 591–619.

Byrne, David P., Nicholas de Roos, A. Rachel Grinberg, and Leslie M. Marx (2025). "The Role of Platforms for Facilitating Anticompetitive Communication". *Cartels Diagnosed: New Insights on Collusion*. Joseph E. Harrington, Jr. and Maarten Pieter Schinkel, eds. Cambridge University Press.

Chassang, Sylvain, Kei Kawai, Jun Nakabayashi, and Juan Ortner (2022). "Robust Screens for Non-Competitive Bidding in Procurement Auctions". *Econometrica* 90, 315–346.

Chen, Gang, Niels G. M. Rytter, Liping Jiang, Peter Nielsen, and Lars Jensen (2017). "Pre-announcements of Price Increase Intentions in Liner Shipping Spot Markets". *Transportation Research Part A* 95, 109–125.

Cramton, Peter and Jesse A. Schwartz (2000). "Collusive Bidding: Lessons from the FCC Spectrum Auction". *Journal of Regulatory Economics* 17, 229–252.

Detecting Cartels for Ex Officio Investigations (2024). OECD Roundtables on Competition Policy Papers. OECD.

Duso, Tomaso, Joseph E. Harrington, Jr., Carl Kreuzberg, and Geza Sapi (July 2025). *Public Communication and Collusion: New Screening Tools for Competition Authorities*. DIW Discussion Papers 2131. DIW Berlin.

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

Harrington, Joseph E., Jr. (2008). "Detecting Cartels". *Handbook of Antitrust Economics*. Paolo Buccirossi, ed. The MIT Press.

— (2022). "Collusion in Plain Sight: Firms' Use of Public Announcements to Restrain Competition". *Antitrust Law Journal* 84, 521–564.

— (2025). "Collusion with Non-express Communication: Retail Gasoline in Norway". *Cartels Diagnosed: New Insights on Collusion*. Joseph E. Harrington, Jr. and Maarten Pieter Schinkel, eds. Cambridge University Press.

Harrington, Joseph E., Jr. and David Imhof (2022). "Cartel Screening and Machine Learning". *Stanford Computational Antitrust* 2, 133–154.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (2019). "Firm-Level Political Risk: Measurement and Effects". *The Quarterly Journal of Economics* 134.4, 2135–2202. DOI: 10.1093/qje/qjz021. URL: https://doi.org/10.1093/qje/qjz021.

Hoberg, Gerard and Gordon M. Phillips (2016). "Text-Based Network Industries and Endogenous Product Differentiation". *Journal of Political Economy* 124.5, 1423–1465. DOI: 10.1086/688176. URL: https://doi.org/10.1086/688176.

Huber, Martin and David Imhof (2019). "Machine Learning with Screens for Detecting Bid-rigging Cartels". *International Journal of Industrial Organization* 65, 277–301.

In the Matter of U-Haul International, Inc. and AMERCO (2010). 150 F.T.C. 1. Federal Trade Commission. URL: https://www.ftc.gov/legal-library/browse/cases-proceedings/081-0157-u-haul-international-inc-amerco-matter.

Kawai, Kei and Jun Nakabayashi (2022). "Detecting Large-Scale Collusion in Procurement Auctions". *Journal of Political Economy* 130, 1364–1411.

Kawai, Kei, Jun Nakabayashi, Juan Ortner, and Sylvain Chassang (2023). "Using Bid Rotation and Incumbency Patterns to Detect Collusion". *Review of Economic Studies* 90, 376–403.

Kepler, John D. (2021). "Private Communication among Competitors and Public Disclosure". *Journal of Accounting and Economics* 71, 101387.

Korinek, Anton (2023). "Generative AI for Economic Research: Use Cases and Implications for Economists". *Journal of Economic Literature* 61.4, 1281–1317. DOI: 10.1257/jel.20231736. URL: https://doi.org/10.1257/jel.20231736.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan (2025). *Large Language Models: An Applied Econometric Framework*. NBER Working Paper 33344. National Bureau of Economic Research. DOI: 10.3386/w33344. URL: https://doi.org/10.3386/w33344.

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe (2022). "Training Language Models to Follow Instructions with Human Feedback". *arXiv preprint arXiv:2203.02155*. DOI: 10.48550/arXiv.2203.02155. URL: https://doi.org/10.48550/arXiv.2203.02155.

Pawliczek, Andrea, A. Nicole Skinner, and Sarah L. C. Zechman (2022). "Facilitating Tacit Collusion Through Voluntary Disclosure: Evidence from Common Ownership". *Journal of Accounting Research* 60, 1651–1693.

Sheng, Qiwen and Tomislav Vukina (2024). "Public Communication as a Mechanism for Collusion in the Broiler Industry". *Review of Industrial Organization* 64, 57–91.

Silveira, Douglas, Lucas B. de Moraes, Eduardo P. S. Fiuza, and Daniel O. Cajueiro (2023). "Who Are You? Cartel Detection Using Unlabeled Data". *International Journal of Industrial Organization* 88, 102931.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). "Attention Is All You Need". *Advances in Neural Information Processing Systems 30*. Isabelle Guyon, Ulrike Von Luxburg, Samy Bengio, Hanna Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, eds. Curran Associates, Inc., 5998–6008. URL: https://papers.neurips.cc/paper/7181-attention-is-all-you-need.

Appendix

A Performance of Alternative Prompts

We assess LLM responses on the benchmark sample against the human ratings. Table A.1 reports results using only the first query for each prompt-transcript pair, while Table A.2 reports results using the average across the initial query and 10 repeated queries for each prompt-transcript pair.

For benchmarking we report the following metrics based on the corresponding confusion matrix for each prompt:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP}, \quad \text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Table A.1: Prompt Benchmarking Against Human Ratings (First LLM Run)

Prompt	TP	FP	TN	FN	Precision	Recall	Specificity	F1
Basic Variants								
SimpleScoreV1	3	0	254	132	1.000	0.022	1.000	0.043
SimpleFlagV1	9	3	251	126	0.750	0.067	0.988	0.122
SimpleFlagScoreV1 (Score)	19	5	249	116	0.792	0.141	0.980	0.239
SimpleFlagScoreV1 (Flag)	22	11	243	113	0.667	0.163	0.957	0.262
Refined Variants								
PriceCapacityV1	27	7	247	108	0.794	0.200	0.972	0.320
PriceCapacityV2	26	8	246	109	0.765	0.193	0.969	0.308
PriceCapacityV3	60	16	238	75	0.789	0.444	0.937	0.569
PriceCapacityV4	84	53	201	51	0.613	0.622	0.791	0.618
PriceCapacityV5	41	12	242	94	0.774	0.304	0.953	0.436
PriceCapacityV6	44	9	245	91	0.830	0.326	0.965	0.468
SimpleCapacityV8	23	5	249	112	0.821	0.170	0.980	0.282
SimpleCapacityV8.1	34	7	247	101	0.829	0.252	0.972	0.386
SimpleCapacityV8.1.1	42	8	246	93	0.840	0.311	0.969	0.454
SimpleCapacityV8.2	33	10	244	102	0.767	0.244	0.961	0.371
SimpleCapacityV8.3	33	6	248	102	0.846	0.244	0.976	0.379
SimpleCapacityV8.4	32	8	246	103	0.800	0.237	0.969	0.366

Notes: Reported metrics use the Aryal et al. (2022) manual binary indicator for capacity discipline, our manual score rating for U-Haul transcripts, binarized at 50, and the LLM response score, binarized at 75. Where the LLM output includes both a score and a binary flag, we report the metrics on separate rows. All prompts are run using the gpt-4o-mini model on all benchmark sample transcripts.

Table A.2: Prompt Benchmarking Against Human Ratings (Average of 11 LLM Runs)

Prompt	TP	FP	TN	FN	Precision	Recall	Specificity	F1
Basic Variants								
SimpleScoreV1	0	0	254	135	0.000	0.000	1.000	0.000
SimpleFlagV1	2	0	245	111	1.000	0.018	1.000	0.035
SimpleFlagScoreV1 (Score)	2	1	253	133	0.667	0.015	0.996	0.029
SimpleFlagScoreV1 (Flag)	4	2	237	87	0.667	0.044	0.992	0.082
Refined Variants								
PriceCapacityV1	13	1	253	122	0.929	0.096	0.996	0.174
PriceCapacityV2	17	1	253	118	0.944	0.126	0.996	0.222
PriceCapacityV3	43	11	243	92	0.796	0.319	0.957	0.455
PriceCapacityV4	77	25	229	58	0.755	0.570	0.902	0.650
PriceCapacityV5	19	3	251	116	0.864	0.141	0.988	0.242
PriceCapacityV6	20	2	252	115	0.909	0.148	0.992	0.255
SimpleCapacityV8	7	1	253	128	0.875	0.052	0.996	0.098
SimpleCapacityV8.1	12	0	254	123	1.000	0.089	1.000	0.163
SimpleCapacityV8.1.1	17	1	253	118	0.944	0.126	0.996	0.222
SimpleCapacityV8.2	11	0	254	124	1.000	0.081	1.000	0.151
SimpleCapacityV8.3	12	0	254	123	1.000	0.089	1.000	0.163
SimpleCapacityV8.4	10	0	254	125	1.000	0.074	1.000	0.138

Notes: Reported metrics use the Aryal et al. (2022) manual binary indicator for capacity discipline, our manual score rating for U-Haul transcripts, binarized at 50, and the average LLM response score across the initial run and 10 repeated queries, binarized at 75. Where the LLM output includes both a score and a binary flag, we report the metrics on separate rows, and assign the flag value using a threshold of 0.5 for the averaged flag from the LLM responses. All prompts are run using the gpt-4o-mini model on all benchmark sample transcripts.

B Human Audit Sample Assessment

The transcript-level human-audit assessment file used in this project is available at this link. It includes the audited transcript identifiers, manual assessments, and supporting coding notes.

C Companies with Collusive Content from the Human Audit (Number = 71)

Table C.1: Companies with Collusive Content from the Human Audit

Company	**Primary Market Supplied**
Air France-KLM	Passenger air transportation
AirTran	Passenger air transportation
Alliance Resource Partners	Thermal coal (energy coal)
American Airlines	Passenger air transportation
Ampco-Pittsburgh	Forged and cast metal components
Atwood Oceanics	Offshore oil drilling services
Avis Budget	Car rental services
Bharti Airtel	Mobile and fixed telecommunications
Boral	Construction materials (cement, aggregates)
Cameco	Uranium mining and supply
Casella Waste	Solid waste collection and disposal
Cementir	Cement production
CEMEX	Cement and ready-mix concrete
Charles River Laboratories	Preclinical research services for pharmaceuticals
Cleveland-Cliffs	Iron ore and steel production
CNX Gas	Natural gas production
Coca-Cola Bottlers Japan	Non-alcoholic beverages (soft drinks)
Companhia Siderúrgica Nacional	Steel production
Continental Resources	Oil and gas exploration and production
Corning	Specialty glass and materials
Daqo New Energy	Polysilicon for solar panels
Delta Airlines	Passenger air transportation
Embraer	Commercial and regional aircraft manufacturing
Essity	Hygiene and tissue paper products
Expand Energy	Oil and gas exploration and production
Ferroglobe	Silicon metal and manganese alloys
Frontier	Passenger air transportation
Glatfelter	Specialty paper products

Company	**Primary Market Supplied**
Glencore	Commodity trading and mining
Goldcorp	Gold mining
Gol Linhas	Passenger air transportation
Green Plains	Ethanol and biofuels
Gulfport Energy	Natural gas production
Harley-Davidson	Motorcycles
Huntsman	Specialty chemicals
ICL	Fertilizers and specialty minerals
Italcementi	Cement production
JinkoSolar	Solar panels
Kirby	Inland marine transportation (tank barges)
Micron Technology	Memory semiconductors
Norske Skogindustrier	Newsprint and publication paper
O-I Glass	Glass containers
Oladapco	Oil products distribution (Nigeria)
Ovintiv	Oil and gas exploration and production
PGS	Seismic data services for oil and gas
Pilgrim's Pride	Poultry products
ProFrac Holding	Hydraulic fracturing services
Prosafe	Offshore accommodation rigs
PT XLSMART Telecom	Mobile telecommunications
Reliance Communications	Telecommunications
Sappi	Paper and pulp
Seadrill	Offshore oil drilling
Seagate Technology	Data storage (hard disk drives)
Six Flags Entertainment	Amusement and theme parks
Smithfield Foods	Pork and processed meat products
STEP Energy	Oilfield services
Superior Energy	Oilfield services
Suzano	Pulp production
Telecom Italia	Telecommunications
Teva Pharmaceuticals	Generic pharmaceuticals
Tidewater	Offshore marine services
Trip.com Group	Online travel booking platforms
Ultra Petroleum	Natural gas production
United Airlines	Passenger air transportation
US Airways	Passenger air transportation
Vodafone	Mobile telecommunications
Waste Management	Solid waste collection and disposal
Western Digital	Data storage devices

Company	Primary Market Supplied
WestRock Paper and Packaging	Paperboard and packaging
WH Group	Pork and meat products
Wienerberger	Clay bricks and building materials

Notes: For the 301 transcripts with the highest LLM-validated scores used in the human audit, these are the companies whose transcripts were found to have collusive content. The "primary market supplied" was generated by ChatGPT 5.2 and then verified by research assistant Vivian Zhang.

D Categorization of Collusive Content

Industry reduced supply or capacity and references competitors' conduct or the industry's interests in a manner supportive of such conduct (40 transcripts)
508031, 597966, 505840, 508027, 419393, 352683, 508036, 352685, 392548, 481676, 481654, 602956, 419394, 481658, 481659, 174683, 416422, 524742, 392547, 481690, 427773, 591146, 597969, 259613, 363812, 381391, 185218, 389910, 634653, 260640, 346016, 611671, 569063, 261028, 348864, 340019, 508199, 215550, 190662, 445956

Industry needs to reduce supply or capacity or show discipline (30 transcripts)
183145, 418022, 267433, 244726, 363814, 519067, 174682, 427772, 627577, 406151, 481652, 325885, 260639, 611639, 494779, 508184, 381486, 617313, 352688, 342599, 199854, 227682, 399246, 260628, 199856, 436918, 260643, 228511, 327069, 617015

The industry has shown or will show supply or capacity or price discipline (12 transcripts)
196851, 392562, 227685, 399267, 396666, 445959, 508185, 264077, 309604, 558080, 396668, 431603

Industry needs to raise prices (12 transcripts)
250287, 408614, 215003, 348098, 558083, 558091, 321896, 300093, 375388, 363809, 305524, 324730

The firm is being or will be less aggressive in pricing and references competitors' conduct in a manner encouraging such conduct (8 transcripts)
558086, 377924, 375653, 309638, 223177, 526439, 631460, 321901

Miscellaneous (7 transcripts)
States an industry objective that will tend to lead to higher prices [325064, 408622]
Industry will limit capacity or supply or not overproduce [252937, 445983]
All firms raised prices and pricing strength is expected to continue [257344]
Firm welcomes consolidation, whether it's done by us or others as it will add to discipline and make a better market for all of us [360150]
Firm will not oversupply and hopes its dealers stop competing with themselves [548932]

E Content Analysis of False Positives

Out of the 301 LLM-validated transcripts with the highest scores, 192 were classified by the human auditor as false positives. Of those, 29 of them lacked any relevant content and have been interpreted as hallucinations. The other 163 contain some content which, in principle, could be part of a collusive message. Specifically, the transcripts refer to the firm's competitive conduct or the state of market competition though did not connect it to the conduct of competitors or the industry and thereby fell short of facilitating coordinated conduct. Here we describe the most common content that appeared in the LLM's excerpts and appeared to be the reason for the LLM assigning it a high score.

67 transcripts referred to a firm reducing, restricting, or managing its supply or capacity in a manner consistent with higher prices or avoiding lower prices. However, it did not say or suggest that this conduct is something the industry should or will or needs to do.

> "Our intention [is] to maintain a disciplined sales approach. Rather than selling at material decrease prices. We implemented a strategy to hold back product [and] resulting in a significant buildup of the inventories."
> [OCI N.V. | Earnings Calls | 24 May 2019]

> "We will continue to manage our customer needs and will not restart capacity on a whim just to add tonnage to the spot market. That would not be good for anyone. Value over volume is a simple philosophy that we will carry on into the future, and it's why you are not going to hear me talk about capacity utilization or even market share."
> [Cleveland-Cliffs Inc. | Earnings Calls | 25 Feb 2021]

> "We stand by our strategy of not forsaking value for volume and protecting our profitability with continued actions on our cost structure [and we] cut our production capacity."
> [PIRELLI & C. Società per Azioni | Earnings Calls | 12 Nov 2012]

> "We go down by elevator and we go up by stairs. So to avoid this huge volatility, we are aiming ... to reduce capacity."
> [Fibria Celulose S.A. | Earnings Calls | 26 Apr 2017]

By noting its leadership position or forecasting constrained industry supply, some announcements do get close to suggesting that firms coordinate on reducing supply or capacity.

> "Our first step is to stop the significant price declines. The most important tool that we have to accomplish this is to reduce capacity to get glass supply in line with market demand. When you're more than half the market, that really adds up."
> [Corning Incorporated | Analyst/Investor Day | 03 Feb 2012]

"And until now, we have not heard anyone else announcing any capacity enhancements in the short term. None of them is increasing capacity."
[HEG Limited | Earnings Calls | 16 Aug 2021]

"Additional capacity reductions can be expected [and] we forecast an extended period when the ... industry operates a smaller, more profitable footprint. Our outlook is for lumber pricing to gradually increase."
[Conifex Timber Inc. | Earnings Calls | 14 Aug 2019]

48 of the transcripts refer to higher prices - whether noting having raised prices or expressing a need to raise prices - or to pricing power. Such a statement might be made in the context of reducing supply.

"We reduced our ... production capacity by more than 20% [and] we have already announced another 20% price hike. We have started to see the sign of improving pricing trend."
[AGC Inc. | Earnings Calls | 06 Aug 2012]

"We're going to... be price stewards. We don't chase volume with price."
[Rayonier Advanced Materials Inc. | Analyst/Investor Day | 10 Oct 2023]

"We have taken a call ... to increase prices of cement to help manage the rising cost of production."
[UltraTech Cement Limited | Earnings Calls | 18 Oct 2021]

With a reference to competitors reducing capacity in the context of an emphasis on raising prices, this announcement is getting close to having collusive content.

"In the short term, there's only one priority, and that's price increases and price increase and price increases. We have good capacity utilization. We have strong brands. We always go for pricing every time. Many competitors, especially in the tissue area, announced that they were shutting down paper machines due to this is looked at an Italian player ... and I think we'll see more of that."
[Essity AB (publ) | Company Conference Presentations | 10 Mar 2022]

In several transcripts, a firm announced it was striving to avoid lowering prices.

"We concluded that lowering price to gain share in a market like we are experiencing this year was not the right strategy."
[Avis Budget Group, Inc. | Earnings Calls | 03 Nov 2009]

"We don't want to underprice and overshoot on demand. We keep some of the exclusive capacity in order to be able to sell at high prices. We want to be profitable [even] if we lost a couple of market share percentage points."
[TUI AG | Earnings Calls | 11 May 2022]

"[We are] idling significant chunks of our assets and slashing our participation in weak markets. We're not just going to participate in those weak markets because it's a chase down into the mud."
[Olin Corporation | Earnings Calls | 27 Oct 2023]

A firm indicated a strategy to strengthen pricing power.

"We deliberately allocate inventory to fiscal years and regions to ensure we maintain apparent scarcity and, therefore, protect our brand health and protect our pricing power."
[Treasury Wine Estates Limited | Earnings Calls | 16 Aug 2018]

"We're in the early stages of having a terrific opportunity over the next 5 years to be in a position where we can increase the returns at our disposal facilities, primarily through price. We're one of the only players with a little bit of excess capacity left."
[Casella Waste Systems, Inc. | Company Conference Presentations | 14 Jun 2018]

A firm described how it is being more disciplined and less aggressive which, if it had expressed competitors acting or needing to act in a similar manner, would have encompassed collusive content.

"We do not want to accelerate a rapid decline in price of the products. We will only ramp it up very moderately when we start production because we do not want to accelerate a collapse in price. We want to be more disciplined in our approach."
[LANXESS Aktiengesellschaft | Earnings Calls | 06 Nov 2014]

"We are not chasing market share. We are focusing on quality customers and then driving spend on our network by our existing customers."
[Dialog Axiata PLC | Earnings Calls | 18 Feb 2025]

"We have ... implemented Microchip's disciplined pricing process ... to ensure that we are competitive and don't chase bad business in the pursuit of profitless prosperity."
[Microchip Technology Incorporated | Earnings Calls | 08 Aug 2016]

In this case, the firm is expressing an aspiration for the industry which is getting close to collusive content.

"We would like to drive the industry in a direction where we can move back to more positive margins."
[Total Access Communication Public Company Limited | Earnings Calls | 10 Feb 2015]

A firm noted that it has market power or what the sources of market power are. While in one case it does mention reduced industry capacity being a source, it is of the form of a factual observation.

> "Supply is having a harder time to react to demand [and] that's [where] we're getting the pricing power."
> [Prologis, Inc. | Company Conference Presentations | 10 Sep 2019]

> "When our actual sales pace exceeds our budgeted sales pace, we try to push prices."
> [Hovnanian Enterprises, Inc. | Earnings Calls | 09 Sep 2013]

> "We expect to continue to implement favorable pricing actions ... due to capacity constraints in our industry. We are operating today in a very strong demand environment without enough industry capacity to meet that demand and expect that to continue into the future."
> [Carpenter Technology Corporation | Special Calls | 18 Feb 2025]

There are some transcripts referring to a reduced industry supply or capacity but only in the context of stating it as a fact. Still, the announcements are getting close to content that may encourage competitors to reduce supply or capacity.

> "This moment of rationalization of supply is new in the history of the Brazilian civil aviation industry. We wouldn't like to start or restart price war in the Brazilian market. Our strategy of reducing capacity is indeed in place and it's showing results."
> [Gol Linhas Aéreas Inteligentes S.A. | Earnings Calls | 14 May 2013]

> "[There is] limited new barge construction in the industry and many units going in for maintenance. So we should have a net decline this year in supply. Nobody is really panicking and trying to go out and build any new equipment right now."
> [Kirby Corporation | Earnings Calls | 01 Feb 2024]

In 14 transcripts, the firm refers to industry consolidation by noting that it has been consolidating, needs to consolidate, or will consolidate. In some instances, the firm attributes consolidation to too much supply or capacity or too many firms. In one case, they do seem to be facilitating such consolidation.

> "We expect the industry will consolidate to approximately 5 Pan-India players. Once that happens, the industry will further see tariff hardening, with pricing power returning."
> [Reliance Communications Limited | Earnings Calls | 10 Feb 2014]

> "In order to really improve value creation, we need consolidation and further restructuring. There's a need for further consolidation and restructuring in this industry to be able to really actually take the costs down and balance the markets as well."
> [UPM-Kymmene Oyj | Earnings Calls | 03 Aug 2011]

In summing up, the LLM identified instances in which a firm's message was consistent with less competition in that it referred to reducing supply or capacity or raising or at least not lowering prices. To facilitate collusive conduct, a firm needs to link such content to industry conduct in a manner that encourages similar conduct by competitors. The LLM failed in not verifying that such a link was present in the firm's message. Consequently, the LLM had a high false positive rate.