Organisation for Economic Co-operation and Development

DAF/COMP/GF/WD(2025)9

Unclassified **English - Or. English**

18 November 2025

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

Global Forum on Competition

Artificial Intelligence and Competitive Dynamics in Downstream Markets

- Session I -

1 December 2025

This contribution is submitted by Pakistan under Session I of the Global Forum on Competition to be held to be held on 1-2 December 2025.

More documentation related to this discussion can be found at: oe.cd/aidm.

Mr Antonio Capobianco [Antonio.Capobianco@oecd.org].

JT03576995

Pakistan[*]

Leveraging Artificial Intelligence for detecting anti-competitive activities Pakistan's Journey toward modernized competition enforcement

Table of Contents

[*] Contribution by Dr. Kabir Ahmed Sidhu, Chairman, CCP.

1. Introduction

1. Pakistan's markets are digitizing rapidly, while public procurement, consumer advertising, and capital-market disclosures are producing ever-larger volumes of unstructured information. In this environment, a purely complaint-driven approach cannot deliver timely, economy-wide oversight. The Competition Commission of Pakistan (the **"CCP"**) established the Market Intelligence Unit (the **"MIU"**) in October 2023 to shift the institution from reactive enforcement to a proactive, intelligence-led model that uses artificial intelligence (AI), automation, and data science to detect potential infringements at scale.

2. This working paper, Leveraging Artificial Intelligence for Detecting Anti-Competitive Activities: Pakistan's Journey Toward Modernized Competition Enforcement, has been prepared for presentation at the OECD Global Forum on Competition 2025 in Paris. It documents MIU's first phase of systems development and deployment, showcasing Pakistan's experience in applying AI-based technologies to strengthen competition enforcement. The paper focuses on four complementary pillars that transform scattered public data into structured, enforcement-ready intelligence:

- Public Procurement Automation to screen tenders for collusive patterns;
- an Automated Digital Market Intelligence System to identify deceptive marketing practices across web and social platforms;
- Automated Detection of Merger Cases to flag intended and consummated transactions in listed firms; and
- A Price Monitoring Dashboard with Anomaly Detection to surface unusual pricing dynamics in key commodities.

3. Together, these capabilities expand CCP's line of sight beyond individual complaints to market-wide signals of risk, enabling faster, evidence-based intervention.

4. Methodologically, the systems combine pragmatic automation with rigorous validation. Python/R scrapers, OCR and document AI convert PDFs and webpages into analyzable tables; keyword frameworks and platform APIs generate continuous alerts; large language models help extract complex tabular data from non-uniform reports; and dashboards (R Shiny) translate results into usable workflows for enforcement teams. Each module is tailored to Pakistan's data realities, legacy PDFs, multilingual media, and bandwidth constraints, ensuring that oversight capacity grows sustainably without proportionate increases in staff or cost.

5. The paper is organized as follows. **Section 2** outlines the establishment, mandate, and operational performance of the Market Intelligence Unit (MIU), highlighting its transition toward proactive, intelligence-based competition enforcement. **Section 3** details the *Public Procurement Automation* framework, describing the complete *Collusive Bidding Detection Mechanism (CBDM)*, from large-scale tender downloading and OCR-based data extraction to rule-based collusion screening. **Section 4** presents the *Automated Digital Market Intelligence System*, which integrates Google Alerts, Talkwalker, Social Searcher, and the Meta Ad Library for detecting deceptive marketing practices across digital platforms. **Section 5** explains the *Automated Detection of Merger Cases*, encompassing PSX announcement scraping, AI-driven annual-report data extraction, and visualization through an R Shiny dashboard. **Section 6** introduces the *Price Monitoring Dashboard with Anomaly Detection*, illustrating the application of econometric indicator-saturation techniques to identify abnormal pricing behavior. **Section 7** concludes with policy

implications and recommendations for scaling AI-enabled enforcement within Pakistan's competition framework.

2. Market Intelligence Unit

6. The MIU was established on 23 October 2023 as a separate department within the CCP. Its creation marked a major shift from reactive to proactive enforcement. Previously, the CCP primarily responded to news reports, public complaints, or government referrals regarding anti-competitive conduct. With the establishment of MIU, the Commission adopted a data-driven approach focused on proactive detection, greater efficiency, and improved market stability. The MIU primarily identifies breaches of the Competition Act, 2010 (the '**Act**') and refers them to relevant departments for investigation and enforcement.

7. The CCP has inducted quality human resources equipped with relevant qualifications and experience which include a team of Economists, Econometricians, and Data Analysts.

2.1. Scope and Functions

8. The MIU is tasked with proactively detecting violations of Chapter II of the Act, including abuse of dominance, cartelization, prohibited agreements, deceptive marketing, and merger transactions, through modern detection tools. Its function includes:

- Detection and provisioning of reports indicating violation(s) of the Act to respective departments.
- Development of database for structural and behavioral analysis of different sectors of the economy.
- Finding avenues for information collection to gather evidence regarding anti-competitive activities.

2.2. MIU's Approach

- **Automation:** Use of Machine Learning/AI Model/Data Sciences for proactive detection of anti-competitive practices.
- **Use of Economics:** Market screening [analysis of structural and behavioral screens] and Collusion factors.
- **Industry Monitoring:** Open-source monitoring, infiltration and surveys.

2.3. MIU's Strategic Focus

- Effect on public at large (i.e., essential commodities, real estate/construction, transport, education and health).
- Prone to cartelization due to market structure and history of collusions.
- Sectoral weightage in GDP.

2.4. MIU Performance

9. The performance of the MIU mainly revolves around its detection activities and special initiatives for the implementation of AI based detection tools. The details are provided below.

2.4.1. Detection Activities

10. Since its inception, the MIU has identified more than 200 cases related to deceptive marketing, mergers control, cartelization, and prohibited agreements with an overall efficiency rate of 92%. Actions taken by other departments of CCP since MIU leads are provided below:

Departments	MIU Leads
Office of Fair-Trade Department	124
Merger & Acquisition Department	58
Cartel & Trade Abuse Department	25
Exemption Department	05
Total	**212**

2.4.2. Special Initiatives

11. The transformation from manual detection to AI-driven automated systems is also underway. In this regard following actions have already been taken:

- **(a)** Development of an AI-driven web-scraping system for the Public Procurement Database (**"Public Procurement Automation"**).
- **(b)** Strengthening real-time oversight of online commercial activities through continuous ad monitoring, and data analysis (**"Automated Digital Market Intelligence System")**.
- **(c)** Enhancing merger oversight by analyzing historical shareholding data to identify transactions meeting regulatory thresholds ("**Automated Detection of Merger Cases**"**)**.
- **(d)** Improving market transparency through real-time price tracking and anomaly detection across key commodities (**"Price Monitoring Dashboard with Anomaly Detection")**.

3. Public Procurement Automation

12. Pakistan's public procurement system is vast, complex, and structurally exposed to collusive risks. Each year, thousands of tenders are issued by federal ministries, provincial departments, state-owned enterprises, and other public sector bodies. The Competition Commission of Pakistan's (CCP) Cartel and Trade Abuse Department has already carried out several investigations uncovering such anti-competitive practices in recent years which indicates that public procurement carries risk of cartelization.

13. Historically, procurement data in Pakistan has primarily existed in physical form, which made systematic monitoring and analysis extremely difficult. To address this, the Public Procurement Regulatory Authority (the **"PPRA"**) launched the E-Pak Acquisition & Disposal System (E-PAD) as a pilot project in March 2023, which was later officially

rolled out in the Fiscal Year 2024. This initiative marked a major milestone toward digitizing the tendering process and improving transparency. However, despite this progress, a significant volume of historical data remains in unstructured, scanned, and OCR-based formats, creating a major challenge for detecting collusive bidding patterns. Before E-PAD, the CCP had to manually gather and review tender data from various procuring agencies and authorities, an effort that required substantial time and manpower.

14. To overcome these limitations, the MIU of the CCP has made a pioneering contribution by developing Artificial Intelligence (AI) and Machine Learning (ML) tools capable of processing and analyzing this vast, unstructured dataset. These technologies automatically extract, structure, and detect suspicious patterns in tender data, converting previously unusable information into actionable intelligence. By transforming a slow manual process into an automated, data-driven system, MIU has significantly enhanced the CCP's ability to proactively detect and investigate anti-competitive conduct in Pakistan's public procurement markets, bridging the gap between legacy data and the country's new era of digital procurement.

15. The system reduced manual human work time from months to hours as depicted in the table below.

Table 1. Manual Review of Burden vs. Automated Capacity

Process	Manual Effort	Automated Benefit
Downloading	53,457 tenders × 1 minute each 891 hours estimated (about 5.5 months of human work)	More than 35,000 tender evaluations downloaded in hours via Python-based scraping
Reviewing	53,457 tenders × 5 minutes each 4,455 hours (about 2.3 years of reading)	OCR plus AI-driven structuring enables near-instant screening and red-flagging

16. This automated system works in following three steps (i) Tender Downloading (ii) Data Extraction and (iii) Collusive Bidding Analysis (herein after collectively referred as (**"Collusive Bidding Detection Mechanism or CBDM"**).

3.1. Tender Downloading

17. The most immediate operational challenge in analyzing the Public Procurement data was not the analysis itself, but the sheer effort required for tender gathering/downloading. For instance, in case of the PPRA only, there were more than 55,000 archived reports, manually selecting, downloading, and subsequently reviewing each file was near to impossible. By that time, it was thought that the whole CBDM process may be automated with help of Artificial Intelligence and Machine Learning. This needs for automation define the core of the data acquisition strategy.

18. The automated data download process, executed through a Python-based scraper, encountered certain bottlenecks in accessing large volumes of public procurement data. The following table outlines the issues faced and the measures taken to address them:

Table 2. Bottlenecks and their Solutions using Python Scrapper

Bottleneck	Description	Solution Achieved by Scrapper
Dynamic Link Discovery	The older archive pages didn't display all report links immediately. The website used dynamic loading, content only appeared after a specific user action (like clicking "next page" or expanding a section).	The scrapper was equipped to go beyond just reading the initial page HTML. It was programmed to mimic the necessary network calls (the direct server requests) that the website made in the background to fetch the full list of report links, ensuring all 57,000 links were correctly identified and cataloged.

Rate Limiting and Server Blocking	Rapidly requesting thousands of large files from a government server is perceived as an attack. The server's defensive mechanisms would quickly trigger rate limits or outright IP blocks, halting the download process entirely.	The scrapper was made "polite" through controlled throttling. A mandatory time delay was inserted between each file download request. This deliberate slowdown prevented the server from being overwhelmed, allowing the scrapper to complete the massive download job steadily and reliably over time without being blocked.
Archival Link Errors	The archive, being old, naturally contained broken links and server timeouts that would cause a basic script to crash, forcing a restart of the entire process.	The scrapper utilized robust error handling logic. When a file request failed (e.g., a 404 error), the script would log the faulty link but automatically bypass it and immediately proceed to the next valid tender file, guaranteeing the process reached full completion without manual intervention.

19. In essence, the Python scrapper transformed a logistical nightmare, the manual retrieval of tens of thousands of files into a manageable, resilient, and throttled automated background task, laying the necessary foundation for the subsequent data extraction first and then the analysis.

3.2. Data Extraction

20. The extraction of structured data from scanned documents requires a multi-stage process involving Optical Character Recognition (OCR), specialized Large Language Models (the **"LLMs"**) for information extraction, and rigorous post-processing. This pipeline converts raw image files into a readily usable structured Microsoft Excel spreadsheets.

21. The process leverages the strengths of both Google Cloud AI (for robust document processing) and Hugging Face models (for flexible, specialized information extraction). The AI tools used in data extraction and challenges faced during process are detailed below:

3.2.1. Step 1 - OCR Data Extraction Process Converting Unstructured Image to Structured Excel

22. The overall data flow involves OCR, specialized model-based extraction, and final data structuring, with challenges addressed at each step.

Table 3. Steps Taken for Data Extraction and Tools Utilized

Step	Description	Tools Utilized (General)
Text Recognition (OCR)	Scanned images of tender reports are fed into a robust OCR engine to digitize the text. The output at this stage is a raw, unstructured text block, often accompanied by coordinate data (bounding boxes) for words and lines.	Google Cloud AI (e.g., Document AI/Vision API) is typically used here for its high accuracy on complex or low-quality scanned documents.
Information Extraction	The raw text and layout information are passed to advanced LLMs. These models are instructed to identify, classify, and isolate specific target data points, such as the *Tender ID*, *Bidder Names*, *Evaluated Bid Price*, and *Awarded Contract Value*.	Hugging Face Models (e.g., specialized document-aware LLMs like LayoutLM, Donut, or fine-tuned open-source models) or Google's Generative AI services are used for this step.
Structuring and Validation	The extracted entities are converted into a uniform, tabular format (e.g., a JSON or Python dictionary). This structured data is then checked against business rules (e.g., ensuring Bid Prices are numerical) and validated for completeness.	Custom Python code leveraging data manipulation libraries, often integrated with Generative AI for final format validation and correction.
Final Output	The fully validated, structured data records are systematically compiled and written to the final usable database format.	Database connectors or libraries for direct export to a structured Excel sheet, ready for final analysis.

3.2.2. Step 2 - Issues and Solutions in OCR Data Extraction

23. The conversion from unstructured OCR output to a reliable, structured database presents specific technical hurdles that require tailored solutions using AI and post-processing logic.

Table 4. Bottlenecks in Extracting Data from OCR files and its Solutions

Bottlenecks	Description	Technical Solution
Low-Quality Scans/Skew	Scanned reports are often distorted, blurry, or have poor contrast. Traditional OCR struggles, leading to character substitution errors (e.g., '1' for 'l', 'S' for '5').	Utilize Google Cloud AI's advanced pre-processing to de-skew and enhance images before OCR. Implement post-OCR Fuzzy Matching/Dictionary Correction using a list of known project names or common terms to auto-correct frequent errors.
Unstructured Document Layout	The target data (Bid Price, Names) can appear in tables, lists, or simple prose, with inconsistent formatting across different procuring agencies.	Use Visual/Layout-Aware Models (e.g., those found on Hugging Face) that use the *visual coordinates* of the text, not just the text itself, to locate fields. This allows the model to correctly identify a field even if its position shifts slightly.
Table Recognition Errors	Complex tables in the tender reports (spanning multiple pages, or merged cells) are often poorly linearized by OCR, mixing up rows and columns.	Leverage Google Cloud Document AI's Form/Table Parsers or fine-tuned specialized models. These tools are trained to explicitly recognize and output table data in a clean row/column structure, preserving the integrity of the bidding records.
Contextual Data Linking	The model may extract a 'Bid Price' but fail to correctly link it to the corresponding 'Bidder Name' when they are visually separated on the page.	Apply a Template-Free Extraction Model (a generative LLM) and instruct it with a clear prompt to output structured JSON. The prompt forces the model to perform the contextual reasoning and link the extracted entities together *before* the data is exported.

3.3. Collusive Bidding Analysis

24. The cleaned and structured tender data was then utilized by the rule-based collusive bidding analysis system, which focused on detecting "Red Flags", abnormal patterns in bidder behavior, pricing, and outcomes in tender over PKR 50 million. The system specifically monitored indicators across three categories (i) Financial/Pricing Behavior (e.g., identical bids, fixed percentage differences, or winning bids significantly exceeding cost estimates), (ii) Participation & Rotation (e.g., bidders taking turns winning contracts or allocating tenders geographically), and (iii) Non-Competitive Behavior (e.g., a winning bidder subcontracting to a losing bidder, or submission of non-genuine "cover bids"). The resulting output was a consolidated Excel dashboard featuring automated red-flag tagging and winner identification.

Table 5. Detailed Collusion Checks (Red Flag Indicators)

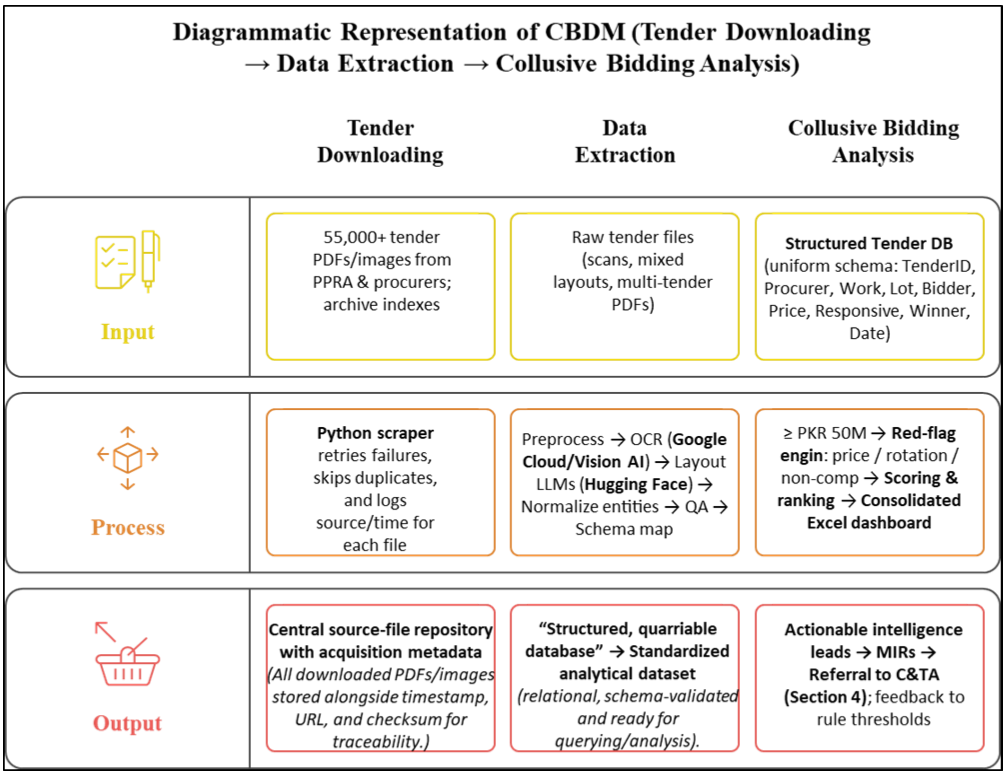
Category	Specific Collusion Red Flags Monitored	Rationale for Suspicion
Financial/Pricing Behavior	**Identical Bid Prices:** Two or more competitors submit exactly the same final bid price. **Exact Percentage Differences:** Bids that are not identical but vary by a fixed, predictable percentage or margin. **High Bid vs. Estimate:** The winning bid is significantly higher than the procuring agency's internal or engineer's cost estimate (signaling a lack of real competition).	Colluders may agree on a **fixed price** or a **fixed differential** to create a facade of competition while ensuring a profitable price for the eventual winner. Prices far above the estimate suggest the bidders are collectively manipulating the market to inflate contract values.
Participation & Rotation	**Bid Rotation/Taking Turns:** A clear pattern where a set of bidders take turns winning similar contracts over time. The winner of the previous bid becomes the high bidder (or does not bid) on the current tender. **Geographic or Customer Allocation:** Firms submit bids only for contracts in specific regions (e.g., Company A only bids in Peshawar, Company B only in Lahore) or for specific clients.	This is classic **bid rotation**, indicating an explicit agreement to divide the market and ensure each member secures a pre-determined share.
Non-Competitive Behavior	**Subcontracting to Losing Bidder:** The winning contractor subcontracts a significant portion of the work to a bidder that submitted a higher price	This suggests the losing bidder was compensated for their non-competitive bid, which

	or was otherwise eliminated. **Sudden Withdrawal:** Regular, expected bidders fail to submit a bid on a tender they would typically pursue. **Non-Genuine/Courtesy Bids:** Losing bids are deliberately incomplete, intentionally high, or contain specific errors (known as **cover pricing**) to ensure a designated competitor wins.	is common in **bid suppression** or **cover pricing** schemes. A sudden failure to bid can indicate an agreement to suppress competition.

25. These automated checks rapidly identified tenders exhibiting statistically significant and behaviorally consistent red flags of potential collusion. The resulting intelligence was consolidated into actionable Market Intelligence Reports (MIRs), out of which **five to six** leads have already been formally transmitted to the Cartel and Trade Abuse Department (C&TA) for detailed investigation and enforcement action.

26. To illustrate the complete analytical pipeline, below presents the Diagrammatic Representation of the **Collusive Bidding Detection Mechanism (CBDM)**, capturing its three sequential phases from Tender Downloading to Data Extraction and Collusive Bidding Analysis. This schematic visualizes how unstructured tender data are systematically transformed into structured evidence, enabling proactive detection and reporting of cartel-like behavior across Pakistan's procurement landscape. A more detailed overview of the above procedure can be seen in the following schematic diagram as well:

Figure 1.



4. Automated Digital Market Intelligence System

27. In its initial phase, the MIU monitored deceptive marketing practices primarily through conventional means, including manual review of print media, digital platforms, and other open-source information. This approach was time-consuming, reactive, and susceptible to oversight.

28. To enhance efficiency and expand monitoring coverage, the MIU introduced a systematic, technology-driven surveillance mechanism for detecting deceptive marketing practices. Under this framework, the MIU conducts sector-wise thematic reviews utilizing tools such as Google Alerts, Talkwalker Alerts, Social Searcher, Meta Ad Library, and customized web-search queries (hereinafter collective referred as “media monitoring tools”) . These tools generate automated alerts based on a list of sector-specific keywords, enabling CCP for a proactive and targeted monitoring.

4.1. Google Alerts

29. The intelligence mechanism’s strength lies in its wide coverage and automated web monitoring. Using Google Alerts, it tracks deceptive marketing keywords, like “clinically proven,” “No.1 whitening cream,” or “Islamabad-based housing projects” across web pages, news, and blogs. Real-time alerts sent to MIU’s email and databases create a continuous, human-free monitoring loop, drastically reducing response time and enabling prompt detection of potential violations under Section 10 of the Act.

4.2. Talkwalker Alerts

30. The system’s core strength is its multilingual, social media–inclusive coverage, surpassing conventional tools like Google Alerts. Using Talkwalker, it monitors not only websites but also real-time content from Twitter/X, YouTube, and online comment sections. This enables detection of deceptive marketing by influencers or small businesses through short-form content often missed by traditional monitoring. Its cross-platform reach significantly expands MIU’s ability to identify false or misleading claims across Pakistan’s diverse digital landscape.

4.3. Social Searcher

31. The tool’s main strength is its advanced social media monitoring enhanced by sentiment and engagement analytics. Using Social Searcher, it gathers public posts from platforms like Facebook, Instagram, X, YouTube, TikTok, and Reddit. Through custom Boolean queries (e.g., “whitening cream” + “FDA approved”), it flags content with potentially deceptive language. This enables MIU to gauge both the spread and impact of such content, helping prioritize enforcement against ads or influencers with the highest risk of consumer deception and market harm.

4.4. Meta Ad Library

32. The tool’s main strength is its function as Facebook and Instagram’s official transparency database. Through the Meta Ad Library, it enables keyword and advertiser searches across all active and inactive ads. It provides verifiable evidence for CCP, including screenshots, advertiser details, funding sources, and key metadata such as target country, dates, and impressions. This allows MIU to directly link deceptive claims to

specific sponsored content, offering credible and admissible evidence for proceedings under Section 10 of the Act.

4.5. Overall Impact of Automation and Digital Surveillance

33. The automation of deceptive marketing detection through multi-platform monitoring has significantly enhanced the Commission's capacity to identify, document, and address violations under Section 10 of the Competition Act 2010. By leveraging tools such as Google Alerts, Talkwalker, Social Searcher, Meta Ad Library, and custom web-scraping scripts, the MIU was able to systematically expand the scope and speed of market surveillance across multiple sectors.

34. As a result of this data-driven approach, MIU successfully generated 64 actionable leads in the beauty and personal-care sector, involving 47 undertakings engaged in potentially deceptive marketing practices. Instead of immediately proceeding toward enforcement, MIU recommended a collective engagement approach, inviting these companies for awareness and guidance sessions on compliance and responsible advertising, thereby advancing CCP's advocacy mandate under Sections 28 and 29 of the Act.

35. In addition, MIU's automated data collection identified over 15 beauty-cream products that concealed the presence of mercury beyond permissible limits in their labeling and promotional material. This initiative attracted national and international media coverage, highlighting the Commission's proactive consumer-protection role in public health and safety.

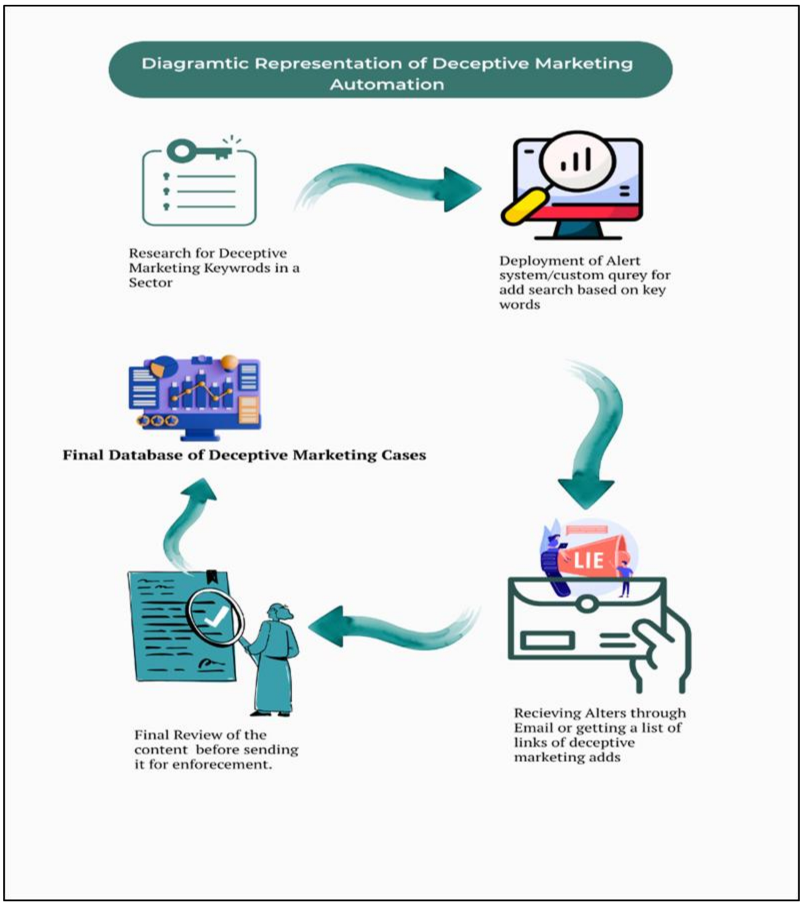
36. Beyond the beauty sector, MIU's sectoral automation framework extended to the real-estate domain, where data from 34 housing societies were compiled and shared with the Office of Fair Trade (OFT). These societies were falsely marketing themselves as located within Islamabad's geographical boundaries, despite being situated outside the capital's administrative limits. The evidence, drawn from digital advertisements, websites, and marketing collateral, has provided OFT with a strong factual basis for potential enforcement under Section 10(2)(b) and (c).

37. Overall, the deployment of automated surveillance systems has transformed MIU's operational model - from manual case-based monitoring to an intelligence-led, preventive, and scalable mechanism. This shift has:

- Enabled early detection and mapping of deceptive patterns across sectors;
- Strengthened institutional coordination between MIU and OFT for evidence-based enforcement;
- Promoted compliance-oriented engagement with industry stakeholders; and
- Enhanced the visibility and credibility of the Commission's work at both national and international levels.

38. Together, these outcomes demonstrate how technology-driven market intelligence can serve as a cornerstone for proactive consumer protection and competitive fairness in Pakistan's evolving digital marketplace. An overview of the above procedure can be seen in the following diagram:

Figure 2.



5. Automated Detection of Merger Cases

39. Historically, the CCP relied primarily on voluntary pre-merger applications submitted by undertakings to identify transactions requiring merger approval under Section 11 of the Act read with Regulation 6 of the Competition (Merger Control) Regulations, 2016 (the "**Merger Regulations**"). However, this approach limited the Commission's visibility to only those mergers voluntarily notified by the parties involved in the Merger Transaction. To expand its detection capacity, CCP began manually monitoring Pakistan Stock Exchange (PSX) announcements and media reports to identify potential ownership changes or acquisitions. Although this manual system was an improvement, it remained resource-intensive, time-consuming, and reactive in nature.

40. To overcome these limitations, the MIU developed a fully automated merger detection framework that integrates artificial intelligence, data scraping, and interactive visualization. This system enables the Commission to identify not only intended mergers but also detect consummated mergers, thereby improving compliance with Pakistan's merger control regime. This system utilized following tools:

5.1. Automation through Google Alerts

41. The MIU established media monitoring tools using predefined keywords such as “acquisition,” “buyout,” “merger,” “stake purchase,” and “ownership change” to automatically track news articles, press releases, and online media coverage of potential merger activities.

42. The system ensures real-time alerts directly to MIU’s shared email account, eliminating the need for daily manual browsing of financial portals or newspapers. This has enabled early detection of domestic and cross-border mergers even before formal regulatory filings are made.

5.2. Python-Based Pakistan Stock Exchange (PSX) Scraper

43. The second automation layer involved the development of a Python-based web scraper designed to systematically download all PSX announcements related to listed companies. Each announcement’s PDF-files is automatically saved in a structured folder system, categorized by company name and date. The scraper also extracts key metadata, such as title, date, and subject keywords, and compiles them into an Excel-based tracking sheet for easy reference.

44. This automated database enables officers to efficiently filter announcements by company, sector, or announcement type (e.g., *Material Information*, *Board Meeting*, *Expression of Interest*). By doing so, it eliminates the need for repetitive manual searches and significantly enhances the speed and accuracy of information retrieval.

45. This automation has significantly reduced the time spent searching for and organizing PSX data, allowing MIU analysts to focus on substantive assessment rather than administrative retrieval.

5.3. Detection of consummated mergers through AI

46. In order to detect consummate mergers in listed companies an AI based system has been developed which works in the following manner:

5.4. Automation of Annual Report Data Extraction

47. Another bottle neck in detection of consummated mergers in listed companies was the systematic extraction of shareholding data from Companies’ financial statements, which are often over 600 pages long and contain scanned or image-based tables with inconsistent formats. Initial attempts using conventional Python and R text-extraction libraries (such as *PyPDF2*, *pdftools*, and *tabulizer*) proved unreliable due to variations in layout and non-text layers.

48. To overcome this, MIU leveraged Large Language Models (LLMs) - specifically Humata.ai and ChatGPT Plus advanced PDF capabilities - to read and interpret these unstructured annual reports. These AI tools were used to:

- Identify and extract the *“Shareholding data”* section from each report;
- Convert extracted tables into structured Excel datasets; and
- Standardize ownership data across multiple years for trend analysis.

49. This innovation transformed static, unsearchable documents into a machine-readable, filterable database, paving the way for automated threshold analysis.

5.4.1. Development of the R Shiny Dashboard

50. The cleaned and structured ownership data were integrated into an interactive R Shiny dashboard developed in-house by the MIU with the following feature:

- Historic shareholding patterns displayed per company and per shareholder;
- Automated comparison of ownership changes across reporting years;
- Real-time detection of potential ownership transfers exceeding the thresholds prescribed under the Competition (Merger Control) Regulations, 2016; and
- Downloadable analytical reports and dynamic filtering options for sectoral or entity-level monitoring.

51. The dashboard thus serves as a live merger intelligence system, providing continuous monitoring and visual analytics to support the Mergers and Acquisitions Department in early identification of potential unnotified transactions.

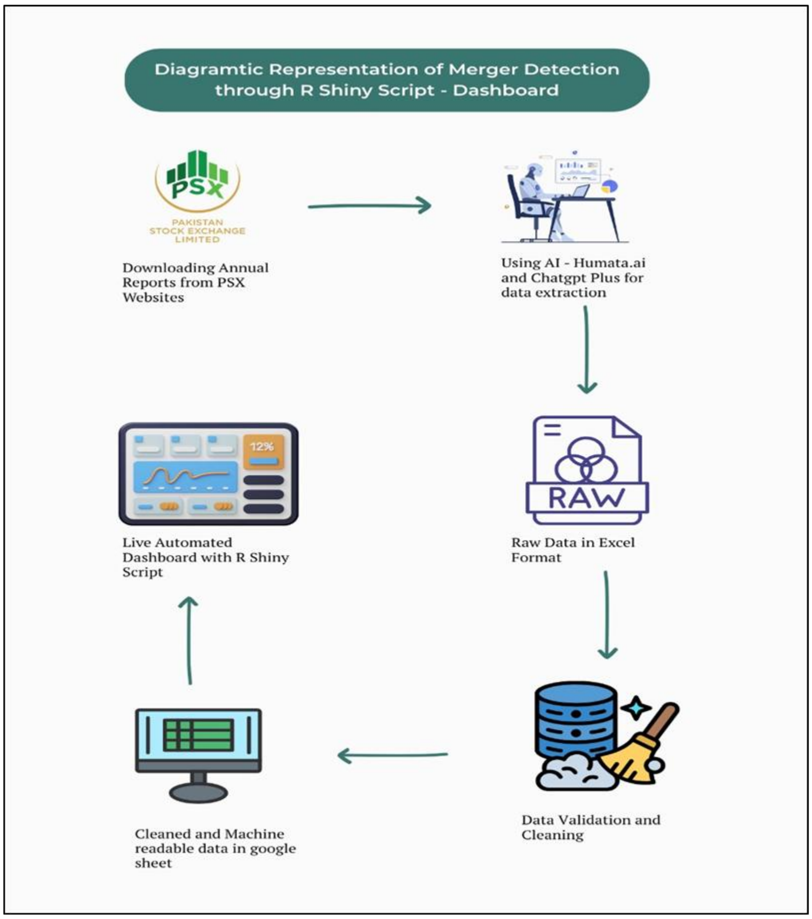
5.5. Overall Impact of Merger Detection Automation

52. The automation of merger detection has transformed CCP’s monitoring process from a manual, reactive approach to a proactive, data-driven system. During the pilot phase, MIU tested the automated framework on a limited dataset of PSX-listed companies and successfully identified 49 potential merger events - 17 through ownership pattern variations and 32 through PSX announcements and media alerts.

53. Each flagged case was analysed to determine whether it qualified for pre-merger notification under Regulation 4(2)(a)–(d) of the Merger Control Regulations, 2016, and whether an application had already been filed with the Commission. Based on this validation, enforcement or formal examination was initiated in over 23 cases within just three months of operation.

54. The system now enables early detection, cross-verification of media and PSX data, and faster enforcement response, significantly reducing the reliance on voluntary filings. This initiative marks a major shift toward intelligence-led regulation, enhancing CCP’s capacity to identify, verify, and act on potential mergers in real time. A detailed overview of the above procedure can be seen in the following schematic diagram:

Figure 3.



6. Price Monitoring Dashboard with Anomaly Detection

55. The development of the Price Monitoring Dashboard (R, Shiny with Anomaly Detection) is a critical initiative aimed at automatically identifying parallel or unusual pricing patterns across Pakistan's commodity markets using Pakistan Bureau of Statistics (the **"PBS"**) data, covering 17 regions and is successfully implemented on cement prices. The whole process is defined by three core phases as discussed below;

- Data Acquisition and Structuring,
- Anomaly Detection, and
- Dashboard Visualization.

6.1. Data Acquisition and Structuring: Taming the Unruly PDFs

56. The first and most formidable challenge lay in data acquisition. The raw material–commodity pricing data from the PBS, was buried within PDF documents, a notoriously unstructured format.

6.1.1. The Challenge: PDF Constraints

57. The data wasn't neatly available in a database; it was locked within tables in static PDFs. Extracting this required not just speed, but accuracy across 17 different regions and various commodities, week after week. The risk was high: a single misalignment during table extraction could cascade, leading to incorrect prices or regions, ultimately undermining the entire anomaly detection process.

6.1.2. The Solution: Scripted Precision

58. This challenge was overcome by creating an automated, two-step preprocessing pipeline using R libraries.

- **Extraction:** Customized scripts were then developed to recognize the typical table layout within the PBS PDFs and extract the raw text and numerical data.
- **Structuring (Tidy Data):** The powerful 'dplyr' package was then employed to rigorously clean, parse, and structure this data. This ensured every price point was correctly associated with its date, region, and commodity, transforming the unruly raw text into a consistent, 'tidy' time series format ready for analysis. The resulting data store was now a reliable source and can be used for the anomaly detection analysis.

6.2. Anomaly Detection: Hunting for the "Unusual"

59. With clean, structured data in hand, the focus shifted to the core intelligence of the project: automatically identifying anomalous pricing behavior.

6.2.1. The Challenge: Finding the Needle in 17 Haystacks

60. In a vast dataset covering multiple regions over long periods, simple threshold-based outlier detection is prone to error. Natural economic fluctuations might be flagged as false positives, or, worse, genuine parallel pricing patterns, where prices across multiple regions move together in an unusual way, might be missed. We needed a statistically robust method that understood the time series nature of the data.

6.2.2. The Solution: Indicator Saturation (Econometric Robustness)

61. The project adopted an advanced data-driven technique *i.e.*, Indicator Saturation. This econometric approach, is specifically designed to detect structural breaks and outliers within a time series model without relying on prior knowledge of when they occurred.

- **Time Series Preparation:** Each region's price data for a specific commodity (like cement) was converted into its own time series object.
- **Model Application:** The Indicator Saturation model systematically introduces dummy variables (indicators) for every data point and then uses a search algorithm to select only the significant indicators.
- **Flagging:** The final output was a Flagged Data set containing only the points confirmed to be statistically significant anomalies. This approach effectively minimized false positives, ensuring that only genuinely unusual pricing events, potentially indicative of anti-competitive behavior, were highlighted.

6.3. Dashboard Visualization

62. The final phase focused on translating complex statistical output into actionable intelligence *via* an intuitive front-end application.

6.3.1. The Challenge: Communicating Complexity Simply

63. The powerful statistical output, model coefficients, p-values, and break dates, is useless to a policy maker or enforcement officer if presented raw. The dashboard needed to clearly show *where* and *when* the anomalies occurred without requiring the user to understand the underlying econometrics.

6.3.2. The Solution: The Shiny Interface

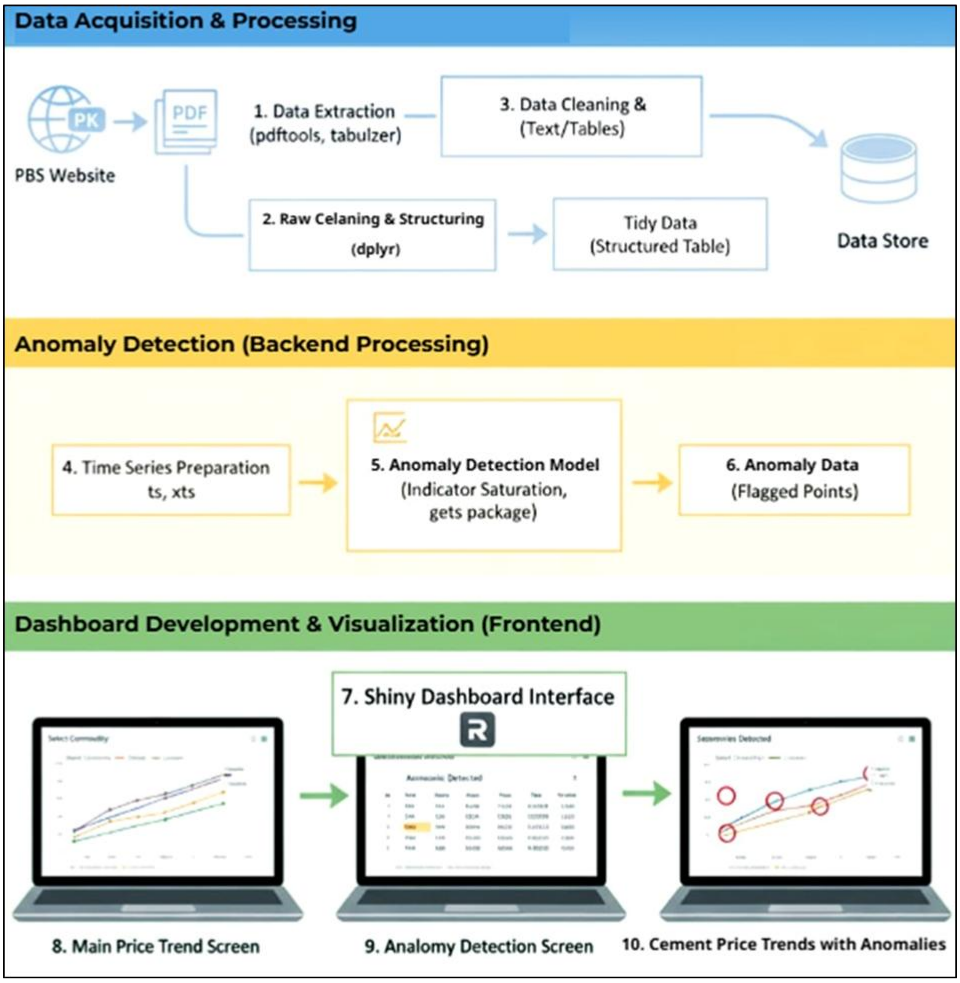
64. The R Shiny framework provided the perfect platform to create a responsive, interactive web application. The dashboard was structured into three highly functional screens:

- **Main Price Trend Screen:** Users could select a commodity (e.g., Cement) and immediately view the aggregate and regional price lines over time. This established the baseline "normal" trend.
- **Anomaly Detection Screen (Table View):** This central screen presented the Anomaly Data in a clean, filterable table using the DT package. It listed the Date, Region, Price, and the Significance of the anomaly, allowing enforcers to quickly triage and prioritize investigation based on the most statistically confirmed breaks.
- **Visual Confirmation Plot:** Crucially, this screen used interactive charting (**plotly**) to superimpose the flagged anomalies directly onto the historical price lines. By seeing the red circle marking the anomaly on the time series plot, the enforcement team could instantly confirm the unusual price movement in context.

65. The resulting dashboard provided a single source of truth, a powerful tool that successfully translated complex data extraction and advanced anomaly detection into clear, actionable intelligence, supporting more effective competition enforcement in the markets under scrutiny.

66. A more detailed overview of the above procedure can be seen in the following schematic diagram as well:

Figure 4.


Data Acquisition & Processing
PK
PDF
PBS Website
1. Data Extraction
(pdftools, tabulzer)
3. Data Cleaning &
(Text/Tables)
Data Store
2. Raw Celaning & Structuring
(dplyr)
Tidy Data
(Structured Table)
Anomaly Detection (Backend Processing)
4. Time Series Preparation
ts, xts
5. Anomaly Detection Model
(Indicator Saturation,
gets package)
6. Anomaly Data
(Flagged Points)
Dashboard Development & Visualization (Frontend)
7. Shiny Dashboard Interface
8. Main Price Trend Screen
9. Analomy Detection Screen
10. Cement Price Trends with Anomalies

7. Conclusion

67. This paper demonstrates that a competition authority does not need perfect data or costly enterprise systems to build proactive, intelligence-led enforcement. By combining targeted automation (scrapers, OCR, layout-aware LLMs), disciplined data engineering (schema mapping, fuzzy resolution, Quality Assurance), and purpose-built analytics (rule-based collusion screens, indicator-saturation anomaly detection), the MIU has turned fragmented public records into structured, enforcement-ready evidence for cartel screening, deceptive-marketing control, and merger oversight.

68. The results are tangible. In public procurement, the end-to-end CBDM reduced years of manual review to days and has already generated five to six actionable cartel leads formally transmitted to C&TA for investigation. In consumer markets, the automated digital intelligence stack produced sector-wide coverage and prioritized cases under Section 10. In merger control, PSX automation and LLM-assisted extraction created a live dashboard that surfaces potential notifiable or consummated transactions earlier in the cycle. The price-monitoring module translates complex econometrics into clear, traceable signals for enforcement teams. Collectively, these tools expand CCP's line of sight, shifting the institution from reactive complaint handling to front-foot market surveillance.

69. Equally important is the governance model. MIU's approach embeds human review, evidentiary traceability, and due-process safeguards at each stage, logging provenance, preserving source documents, and treating model outputs as leads, not findings. This balances innovation with legal robustness, ensuring that automation strengthens, not substitutes, investigative judgment.

70. For jurisdictions facing high procurement spend, concentrated suppliers, fast-moving digital advertising, and limited staff capacity, Pakistan's experience offers a replicable, modular blueprint. Authorities can start where data are available, iterate quickly, and scale: (i) integrate E-PAD and procurer portals for real-time tender ingestion; (ii) extend red-flag logic with learning feedback from closed cases; (iii) widen media monitoring to vernacular and short-form content; and (iv) institutionalize dashboards and M&E metrics to track deterrence and outcomes.

71. The broader lesson is straightforward: AI changes what the authority can see. By upgrading visibility into procurement, advertising, and ownership structures, and converting that visibility into timely, credible leads, MIU materially improves deterrence, evidentiary quality, and response speed. This is modernization on a realistic budget: a practical path by which competition agencies can protect consumers and competitive neutrality, even in data-scarce environments.