STANFORD
COMPUTATIONAL
ANTITRUST

IMPLEMENTATION SURVEY

# Computational Antitrust Within Agencies: 3rd Annual Report

Thibault Schrepel* & Teodora Groza** (editors)

**Abstract**. In the first quarter of 2024, the Stanford Computational Antitrust project team invited its partnering antitrust agencies to share their advances in implementing computational tools. The 16 contributions received provide a distinctly broad geographical representation, offering an overview of global developments. In terms of the substantive focus of the agencies, the main trends that can be discerned are the following: i) emphasis on the detection of bid rigging practices; ii) focus on practices that harm end consumers (e.g., price increases in fuels, airplane tickets; detection of dark patterns); and iii) significant investments in detecting the public perception of market competition. In terms of the computational tools used, the key developments are: i) the gradual integration of large language models (LLMs) in the daily operations of the agencies; ii) increasing reliance on machine learning (ML) tools for analyzing sizeable volumes of textual data; and iii) the development of Proof-of-Concept Application Programming Interfaces (APIs).

* Associate Professor at the Vrije Universiteit Amsterdam; Faculty Affiliate at Stanford University CodeX Center (creator of the Stanford Computational Antitrust project); Invited Professor at Sciences Po Paris & University of Paris 1 Panthéon-Sorbonne.

** Ph.D. student at Sciences Po Paris and Editor-in-Chief of the Stanford Computational Antitrust Journal.

## Table of Contents

## Brazil
## Administrative Council for Economic Defense (CADE)

### I. Context

The Brazilian Competition Authority has considerable experience in the development of data mining and screening techniques to enhance the detection of signs of bid-rigging. The Cerebro project harnesses a wide array of computational tools to detect and investigate cartels operating in public procurement processes, capitalizing on Brazil's robust culture of public data accessibility.

One example of this culture are public datasets: Comprasnet, the Brazilian Federal Government's Procurement Portal launched in 1997, electronically publishes bidding notices, calls for bids (of all modalities), and summaries of contracts signed by the Federal Public Administration. This movement towards transparency continuously grew over the last years and recently was enforced by the enactment of a new public procurement law in 2021. The new public procurement law mandated that all Brazilian procurements be online (with few exceptions). Subsequently and related to Brazilian public data culture, the Portal Nacional de Compras Públicas (PNCP) was set up to spread comprehensive information about tenders, bidders, bids, winners, reference prices, and related data, including all states and municipalities.

Given this extensive range of public procurement data, the first challenge lies in accessing and processing these datasets to facilitate Cerebro's primary aim: the detection of collusive behavior. In this context, the data must be selected and processed as close as possible to the needs of Cade’s anticartel core business, which is intrinsically related to the active search of new cases on one hand, and the enrichment of opened cases with relevant data on the other hand. This work is done alongside other units of the General Superintendence.

### II. Techniques and Computational Tools

A significant challenge for Cerebro is developing methodologies and tools that empower case handlers in identifying companies engaged in illicit activities during ongoing investigations. Also, there is a need to help those who are responsible for handling complaints. In this regard, Cerebro’s products can speed up responses to public claims.

The methodologies and tools aid in identifying companies engaged in illicit activities during ongoing investigations. As a result, it is crucial to propose indicators or markers that simplify the detection of patterns typically associated with collusive behavior.

The information employed by Cerebro in antitrust investigations can be categorized as follows:

1. Information focused on firms: Encompasses details about companies and their systems, including office locations and the IP addresses used during participation in electronic tenders, for example.
2. Information focused on tenders: Focuses on the behavior of firms within tenders, analyzing participation profiles, including firms frequently taking part together, firms winning tenders but not fulfilling contracts ("rabbits"), firms consistently unsuccessful in tenders, and firms exclusively winning in specific regions.

Both categories are complementary and instrumental in comprehending market behavior. The former finds suspicious coincidences, like shared IP addresses, potentially indicating illicit communication. The latter, coupled with traditional economic screens like irrelevant bids, contributes to the body of evidence suggesting the existence of a cartel.

It is important to note that, as most antitrust agencies, Cerebro lacks access to a comprehensive database with labeled data clearly identifying collusive firms and their corresponding markers. Given this constraint, Cerebro focuses on providing comprehensive information to investigators, who have the ability to discern their relevance within specific case contexts.

The conversion of data into useful information in the context of a case is not straightforward, particularly when there are multiple data and pieces of information that come from other sources. Consequently, Cerebro develops tools that streamline the interpretation of results for non-data scientist experts, facilitating informed decision-making in cases. To this end, Cerebro's team is actively curating a catalog of tools.

These tools can be divided into two generations.

The first generation is built upon techniques like network analytics, statistical, and econometric indicators. Networks elucidate firms' societal compositions, potential shared decision boards, and the organization of tenders and bidders. Traditional indicators, such as variance, cover bidding and structural screens, are currently implemented as Business Intelligence (BI) panels.

The next generation of tools is under development, harnessing modern machine learning techniques, including artificial intelligence (AI) and open-source intelligence. Ongoing research investigates unsupervised learning to address the challenge of limited labeled data. Clusters are defined using various inputs, including bid-rigging screens and communication evidence. Additionally, Cerebro plans to employ natural language processing (NLP) and large language models (LLMs) to enhance the capabilities of procurement document processing, particularly in the search for documents' coincidences and other indirect communication evidence. All these datasets and information can be used as input data to the machine learning model.

To make all the tools mentioned available for the investigators, Cerebro's team must have the ability to build them from the ground up performing:

- **Requirement elicitation:** includes not only talking to case handlers, but also managing cases. The General Superintendence decided on a development tailored to the challenges faced by the investigators.
- **Software development:** Backend (including data security, database operations, API development and dev-ops) and Frontend (User Interface (UI) and User Experience (UX) design).

It is worth mentioning that previously to the implementation of a tool for the investigator's direct use, like a web-app, there is all the research and development that is made using the mainstream data-science toolkit: *jupyter notebooks*, quarto-markdown documents, R and Python scripts, and so on. At the same time, Cerebro's team is also clearly tailored to innovating not only on the techniques itself. So, there are PoCs (proof of concepts) and prototypes of solutions incorporating more performant and secure languages like Rust into the data science workflow.

Cerebro's team is also currently developing a platform, which is a web-app to host the most used tools by the investigators, using the newest architecture and web technologies for the backend, like Docker, Kubernetes, Microservices and Go, alongside more established technologies like Angular for the frontend. The aim is to support the investigators and case handlers from Cade's General Superintendence.

## III. Challenges

Despite all the technical and methodological advances achieved in the last two years, Cade still faces some challenges to reaching its full potential that are worth mentioning. First, the reports produced by Cerebro analysis must be subjected to the scrutiny of professionals who have low or no knowledge about the technologies used, economics, and statistics theory. For example, the judges responsible for authorizing dawn raids could face difficulty understanding and interpreting the data produced. To avoid misinterpretation is necessary for the Cerebro team to clearly communicate the methods, techniques, and results in such a way that is understandable by the average law professional.

Another concern is about the Cerebro project staff. It is necessary to request professionals from other agencies, which is hard, given the scarcity of qualified professionals with the necessary skills in public service. Consequently, selecting and keeping this kind of skilled professional is a great challenge that Cade faces. Furthermore, not only technical skills are needed. The professional must be an analytical person by nature and have soft skills like curiosity and openness to collaborate and engage in interdisciplinary discussions. Those hard and soft skills can be developed, but there is a steep learning curve.

Regarding the demobilization of the data science team risk, Cade encourages employees to participate in long-term training plans and constantly register and manage knowledge of all products developed in the project.

In addition, strengthening partnerships with other Brazilian investigative bodies and other countries' antitrust agencies is also a strategy to face those challenges. The experience of listening and learning from others is enriching, that's why the Cerebro Project is always open for international collaboration, like in the recently established ICN Technologist Group Work.

## Bulgaria
## Bulgarian Commission on Protection of Competition

The Bulgarian Commission on Protection of Competition (CPC, the Commission) currently does not use any sophisticated computational tools for processing internal or external data. While the Commission is trying constantly to improve its efficiency in various areas, there are still a number of challenges created by the introduction of machine learning systems. This is why the Commission is involved in initiatives and projects on the topic and follows the current trends in order to adopt the most optimal system for its needs.

The Commission has a long expertise with bid rigging cases. Taking into account the best practices of other NCAs and international organizations, such as the OECD, the CPC has developed its own Guidelines for fighting bid rigging. The new Guidelines have a list of the most common red flags, indicated by many authorities, most of which are applicable to machine learning and automated search engines.

The CPC has also adopted an internal Methodology for bid rigging screening based on economic and statistical analysis. The screening methodology is applied during the preliminary investigation. The results of the screening can be used for establishing fluctuations in the competitive process during the tender procedures. The analysis aims to compare the winning with the losing bids to detect relations between them. Currently, the CPC is involved in an ongoing project, which aims to make this process automatic with the assistance of a machine learning tool. Moreover, the Commission is in preliminary talks for an AI tool development, based on national needs, which should further enhance its bid rigging detection capabilities.

The Bulgarian Commission enjoys a wide variety of online and easily available databases, updated both by public and private entities. These sources of information allow the CPC to make fast, up-to-date and reliable enquiries, without the need to send formal requests to different bodies. While the Commission relies mainly on information extraction from open source databases, in many cases the options to filter and/or sort the queried data are rich and powerful enough. This makes the manual processing of information sufficient for the ongoing investigation without the need for a designated computational toolkit.

These databases include registries such as:

- **Trade registry** with full access to the company profiles and documents, current Articles of Association, representation rights, links between the shareholders of separate companies, etc.
- **Digital Public Tenders Platform** where all public procurement notices are announced and all corresponding documents are uploaded.
- **Food Prices Information Portal**—a free online comparison tool for basic food prices, such as fruits, vegetables, bread, milk products, meat, rice, etc.
- **Fuel prices tool**—a free online comparison tool for the prices between petrol stations.

Finally, the CPC has recently updated its forensic toolkit (software and hardware), which includes limited, but useful computational tools as well. The Commission is aware that the purchase of contemporary software products is important for its work, however, it also acknowledges that the technical preparation of its staff is even more vital. For this reason, the CPC is actively enquiring other NCAs as well as other investigatory authorities in Bulgaria for the practices and experience with the toolkits available to them.

## Canada

## Canada's Competition Bureau

The Digital Enforcement and Intelligence Branch (DEIB) leverage data and technology to support the Competition Bureau of Canada's enforcement and competition promotion initiatives. DEIB functions as a centre of expertise on digital business practices and technologies and provides specialized skills, including a centralized intelligence function. The branch acts as an early-warning system for potential competition issues at all stages, from initial intelligence to post-investigation. It also gathers intelligence from the marketplace and provides insight on how companies may use technology to compete or to harm competition. The Digital Enforcement Directorate provides analysis and advice on the use of data, technology, and digital tools in markets as well as how data analytics and emerging digital solutions can be integrated into the Bureau's enforcement and promotion work. The directorate also houses the Innovation Station, a space that brings branch experts and Bureau employees together to creatively problem-solve. The Intelligence, Behavioural Insights, and Remedies Directorate helps the Bureau find potentially harmful conduct quickly, understand how decisions are made by consumers and businesses, and ensure consistent and effective resolutions when enforcement action is taken.

DEIB's objective is to provide support and expert advice for colleagues across the organization, to make the organization's work more efficient and effective. The Intelligence Unit works together with our digital enforcement team to collect and analyze data and uncover trends that can be identified to support proactive enforcement. They also work together using statistical techniques and computational methods to proactively detect anti-competitive bidding practices in Canadian markets. The Data and Analytics Unit uses advanced analytics and data science to build tools that create efficiencies and add new capabilities to our investigative and promotional work. This includes projects with a focus on automation, for instance using Natural Language Processing (NLP) techniques with other techniques to automatically identify and measure groups of similar documents for investigations. The Technology Insights Unit focuses on understanding evolving business practices and technologies, along with their impact on competition. The team helps identify emerging threats to competition more quickly and provides expertise on new technologies.

Artificial intelligence (AI) has been a priority area for the Competition Bureau. This year it launched a consultation on how AI may affect competition, in the context of Canadian competition law.[1] The consultation paper provides an overview of the markets involved in AI production, from computational and data inputs, development of AI technologies, and the production of AI products and services.

[1] Canada Competition Bureau, *Artificial intelligence and competition*, https://competition-bureau.canada.ca/how-we-foster-competition/education-and-outreach/artificial-intelligence-and-competition (last accessed June 8, 2024).

## Caribbean Community (CARICOM)
## CARICOM Competition Commission

The CARICOM Competition Commission (Commission) plays a pivotal role in fostering a fair and competitive marketplace within the Caribbean Community (CARICOM) Single Market and Economy (CSME). Established under Chapter VIII of the Revised Treaty of Chaguaramas, the Commission is tasked with promoting competition, preventing anti-competitive business conduct, and ensuring consumer protection across the CSME.

Over recent years, the Commission has significantly advanced its analytical capabilities by integrating various computational tools. These tools, including STATA, SPSS, and Gretl, have enhanced the robustness and precision of the Commission's market studies and competition enforcement activities.

### I. STATA: A Comprehensive Tool for Econometric Analysis

STATA is a powerful software package widely used for data analysis, data management, and graphics. In the context of the Commission's work, STATA's comprehensive suite of econometric tools enables the Commission to conduct detailed statistical analyses and model complex economic relationships.

STATA was instrumental in the Commission's research into whether mergers created efficiency gains in the banking sector in Jamaica. Using Difference-in-Difference and propensity score matching, the study showed that the effects of mergers were statistically insignificant. This finding is important as it disproves the general assumption in the region that mergers inherently create efficiencies. These insights form a crucial part of the Commission's advocacy for deeper relationships with banking sector regulators, promoting a more nuanced understanding of the impacts of mergers.

STATA was also recently utilised to investigate the factors influencing high food prices in Suriname, a country listed by the World Bank in 2023 as one of the top ten countries in the world with the highest levels of food price inflation. The Commission's research examined whether anticompetitive business conduct could be a plausible factor causing the high prices, using the software's built-in ARDL model to control for demand and supply factors. This investigation was crucial in understanding the dynamics behind Suriname's food prices and highlighting

potential areas for regulatory intervention.

## II. SPSS: User-Friendly and Versatile

SPSS, known for its user-friendly interface, is another critical tool in the Commission's arsenal. It is renowned for its statistical analysis, predictive analytics, and data mining capabilities. SPSS is invaluable for conducting market surveys, analysing consumer behaviour, and segmenting markets. The software's versatility allows the Commission to perform basic and advanced statistical procedures, including descriptive statistics, hypothesis testing, and multivariate analysis. By leveraging SPSS, the Commission can glean actionable insights from data, to detect anti-competitive behaviours and assess market dynamics.

The Commission used SPSS in its research to understand the possible factors resulting in airline passengers not receiving refunds during the initial stages of the COVID-19 pandemic. This study was pivotal in identifying the challenges faced by consumers and the systemic issues within the airline industry. The results of this research and the subsequent report provided policy recommendations for the CSME region, advocating for increased consumer protection in airlines. These recommendations aimed to strengthen regulatory frameworks to protect airline customers during significant disruptions better.

The Commission also sought in this Study to identify the primary drivers of airline choice for customers when they travel. Respondents were asked to consider and score the importance of eight (8) attributes that may influence their choice of scheduled airline services. Moreover, the Commission analysed whether the mean scores were consistent when the respondents were grouped into business and leisure travelers, given the debate surrounding whether business and leisure travelers are in the same relevant market. The results showed that ticket prices and past experiences with scheduled airline services appear more important to leisure travelers than business travelers. In contrast, business travelers consider ticket flexibility more important than leisure travelers, suggesting that business and leisure travelers are in separate markets. However, the built-in SPSS independent t-test suggests that small numerical differences in mean scores for leisure and business travelers in the sample are likely by chance, and the two groups view ticket prices as equally important.

## III. Gretl: Open-Source and Accessible

Gretl, an open-source software package, offers a range of econometric tools and time-series analysis capabilities. The Commission favours Gretl because of its accessibility and cost-effectiveness, especially for conducting preliminary analyses. The Commission also prefers using Gretl when seasonally adjusting time series data. This capability is essential for removing seasonal effects from data to reveal underlying trends, making Gretl an invaluable tool for time-sensitive economic analyses. The software supports statistical techniques, including linear regression, instrumental variables, and panel data analysis. The Commission utilises Gretl to perform initial data explorations, validate findings from other software, and educate staff on econometric modelling without the financial burden associated with proprietary software.

## IV. Towards Machine Learning: The Future of Competition Enforcement

Recognising the limitations of traditional econometric tools in detecting complex anti-competitive practices, the Commission is poised to integrate machine learning software into its analytical framework. With its ability to uncover hidden patterns in large datasets, machine learning holds immense potential for revolutionising competition enforcement.

One of the key areas where machine learning can make a substantial impact is in detecting cartels in public procurement. Cartels often use subtle and complex strategies to manipulate bidding processes, making detection challenging with conventional methods. Machine learning algorithms can analyse vast amounts of procurement data, identifying anomalies and patterns indicative of collusive behaviour. For instance, clustering algorithms can group similar bids, while anomaly detection techniques can flag bids that deviate significantly from the norm. By adopting these advanced tools, the Commission can enhance its ability to uncover and combat cartel activities, ensuring fair competition in public procurement.

The shift towards machine learning also reflects the Commission’s commitment to staying at the forefront of technological advancements in competition enforcement and assisting its limited human resources. This transition will involve not only the adoption of new software but also the upskilling of the Commission's workforce, fostering a culture of continuous learning at the Commission.

## V. Conclusion

The Commission's integration of these computational tools has significantly bolstered its capacity to conduct rigorous market studies and enforce competition laws effectively.

The future adoption of machine learning software represents a promising frontier for the Commission. By leveraging the power of machine learning, the Commission can enhance its ability to detect and prevent anti-competitive practices, particularly in public procurement. This forward-thinking approach underscores the Commission's dedication to fostering a competitive, fair, and dynamic market environment within the CSME. As the Commission continues to evolve, its commitment to innovation and excellence will undoubtedly contribute to the sustainable economic development of the Caribbean region.

## Catalonia
## Catalan Competition Authority

The Catalan Competition Authority (ACCO) launched the ERICCA project in 2022 with two main objectives: assisting the organisation's technicians in handling sanctioning files related to Catalan public tenders and using artificial intelligence (AI) to proactively identify public tenders with a higher probability of anti-competitive agreements among participating companies.

One of the initial main challenges of the project was obtaining quality data from Catalan public tenders. This problem was partially mitigated by signing a collaboration agreement with the Department of Economy of the Generalitat de Catalunya.[2] Thanks to this agreement, ACCO can directly access the database used by the Public Contracts Registry. Additionally, data from Dades Obertes is also utilised.[3]

The tool applies an artificial intelligence algorithm to the available data (an unsupervised machine learning model) to classify companies into clusters, taking into account, among other variables, the tenders they have participated in. Subsequently, various statistical parameters are calculated to identify clusters where collusion among the participating companies is more likely. Since mid-2023, the model has been complemented by a neural network.

Between 2023 and early 2024, the project has progressed in five key areas.

### I. Expansion of Analysed Data

In previous versions of ERICCA, the tool did not analyse low-value tenders due to processing capacity limitations.[4] However, in the latest version, thanks to improvements in computational efficiency, the tool also analyses "minor contracts."[5] This enhancement allows the detection of behaviours prohibited by law, such as the artificial splitting of contracts.

[2] Initially, the information was obtained using web scrapping techniques applied to the Public Procurement Services Platform. Contractacio Publica, https://contractaciopublica.cat/en/inici (last accessed June 7, 2024).

[3] Dades Obertes is a portal of the Generalitat de Catalunya (Government of Catalonia) that collects various datasets produced or compiled by public agencies that public administrations make available to citizens so that they can use them freely in a simple and convenient manner.

[4] Under €15,000.

[5] For amounts greater than €3,000.

## II. Facilitation of File Processing

To improve the processing of files, new reports and visualisations have been implemented, offering a clearer and more detailed view of the analysed tenders. These functionalities include:

- **Global Visualisation of Tenders by Contracting Body:** This function shows all tenders issued by a specific body, including the awarded amounts and the average number of competitors by product or service type. This information is crucial for assessing the level of competition and detecting possible irregularities.
- **Scatter Plots to Identify Suspicious Behaviours:** When a contracting body and a product or service type are selected, the tool generates a scatter plot of the bids submitted by all participating companies, represented in different colours. This visualisation allows for the visual identification of behaviours that might indicate collusion among some companies.
- **Analysis of Company Groups:** This functionality shows all possible groups of companies that have participated in tenders, indicating the number of tenders each group has participated in and the awarded amount. Additionally, the user can select the number of companies forming the group (up to a maximum of eight) and view all possible combinations of the companies in the group.
- **Data Export in CSV Format:** To facilitate external analysis and integration with other data processing tools, ERICCA allows downloading all information in CSV format. This function enables technicians to write more detailed reports and use other data processing programmes, such as spreadsheets, for different types of analyses. These advances significantly ease the work of technicians in processing sanctioning files, providing more powerful and intuitive tools to detect and analyse anti-competitive behaviours.

## III. New Statistical Parameters

New statistical parameters have been incorporated to provide more robustness to the tool's results. For example, the tool calculates the expected discount of a tender considering the contracting body, the number of competitors, or the type of product or service tendered and compares it with the obtained discount to identify possible anti-competitive behaviours. The 'random forest' method is also used to flag contracts with unusually low discounts.

## IV. Use of a Neural Network

The proactive detection of possible collusion cases has been complemented with the use of a neural network based on the model developed by researchers David Imhof and Martin Huber.[6] This neural network has learned from data on public tenders from Switzerland, Italy, and Japan, characterising price distributions in competitive and collusive scenarios to predict possible collusion in a given tender with a high level of confidence.[7] The model calculates the bid interaction matrices for pairs of all companies given a set of tenders.

In the example shown below, the variability of the values in each region of this interaction matrix according to the competitive situation is displayed. Visually, clear differences between the two types of behaviours are already observed, but the algorithm can detect these and other more subtle differences to make predictions.

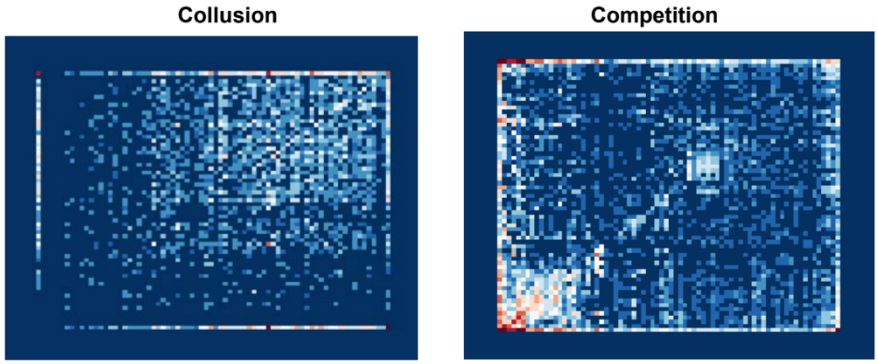


**Figure I:** Collusion and competition

## V. Collaboration Agreements with Other Organisations and Entities

Given the potential of this tool, various institutions have contacted ACCO to use ERICCA. This tool can be used for functions not strictly related to detecting anti-competitive behaviours. As previously mentioned, the tool allows detecting contract splitting, which could violate public procurement regulations. In this regard, the possibility of providing access to some of ERICCA's functions and analyses to other institutions and public bodies is currently being evaluated.

[6] Martin Huber & David Imhof, *Flagging cartel participants with deep learning based on convolutional neural networks* 89 INT. J. INDUS. ORG. 102946 (2023).
[7] In future versions, the model will also learn from Catalan tender data.

The evolution of ERICCA represents a significant step forward in using artificial intelligence to detect and prevent anti-competitive practices in public tenders. This initiative not only improves the efficiency of tender oversight but also reinforces ACCO's commitment to transparency and fair competition in the public market.

## Colombia
## Superintendence of Industry and Commerce

### I. Powers of the Superintendence of Industry and Commerce in Matters of Free Economic Competition

The Superintendence of Industry and Commerce (SIC), through the Deputy Superintendence for the Competition Protection, has the primary function of promoting and protecting free economic competition by exercising *ex ante* and *ex post* functions. On one hand, the Deputy Superintendence conducts *ex ante* or preventive control through the Working Groups of Competition Advocacy, Business Integration, and Compliance Directorate. The first aims to prevent market competition restrictions by reviewing draft administrative acts that regulatory authorities intend to issue. The second is responsible for analyzing business integration operations that may affect free competition. Finally, the Compliance Directorate carries out, among other actions, dissemination, promotion, and training activities related to fostering and building a culture of compliance in matters of free economic competition. Additionally, this Directorate also assists companies in the effective adoption of compliance programs.

In general, *ex ante* control focuses on implementing preventive measures to proliferate a culture of compliance with free economic competition in Colombia. Within the framework of these measures, the SIC seeks to generate awareness among companies and consumers regarding the importance and benefits of adhering to free economic competition policies, as well as ensuring the free participation of companies in the market, consumer welfare, and economic efficiency.

On the other hand, there is *ex post* or corrective control carried out through the Restrictive Competition Practices, Competition Promotion and Protection, and Elite Anti-Collusion Working Groups. This corrective control is undertaken when the Authority has evidenced the possible commission of behaviors that restrict free economic competition in the market. In this sense, these groups are responsible for conducting administrative investigations to determine whether behaviors contrary to the free economic competition protection regime have occurred, such as restrictive competition agreements or abuse of dominant market positions.

## II. Computational Activities Related to the Collection, Preservation, and Analysis of Digital Evidence

The SIC considers it important to highlight the computational activities aimed at collecting, preserving, and analyzing digital evidence related to potential restrictive competition practices. To carry out these activities, specialized tools are used to extract information from devices (computers, emails, cell phones, and/or tablets) identified during unannounced inspections (dawn raids) conducted by the SIC. After collecting the information, the SIC uploads the data into specialized software licensed to access the information and search for words or terms that interrelate the evidence, which may verify potentially anticompetitive behaviors. It is important to note that this software facilitates the visualization of chats, emails, images, among other types of files.

To illustrate the computational process related to the collection, processing, and analysis of various digital evidence, the following elements are presented:

- **Collection:** IT personnel can collect data during unannounced inspections by copying or creating forensic images of the digital data. IT personnel ensure data integrity and, through the chain of custody, the authenticity of the evidence. Additionally, hash values are created to ensure integrity and help verify the authenticity of the copy concerning the original.
- **Processing and Preservation:** After the SIC copies the information, it is processed and subsequently investigated to find evidence. Access to the collected information is restricted, and only the case team can analyze it upon an access request.
- **Analysis:** Following the principal investigator's guidelines and using digital tools, a preliminary filter is applied to the information. The information is filtered by creating tags and identifying important information for the case hypothesis. These results will be subject to further analysis by other members of the investigation team.

## III. Projects Related to Technological Tools that Contribute to the Processing and Analysis of Information

Additionally, the SIC has implemented projects related to technological tools. These projects are aimed at: **(i)** improving the capabilities for detecting potential anticompetitive behaviors, **(ii)** increasing efficiency in examining large volumes of structured and unstructured data, and **(iii)** better allocating human resources.

Currently, these tools are: Sabueso, Inspector, Sherlock, and Búho. As the SIC expands its monitoring capabilities and refines methods for detecting structural and behavioral anomalies, the likelihood of generating timely actions for the protection and promotion of free economic competition increases.

A brief description of the tools that have been developed is provided below.

**1. SHERLOCK**

Developed to reduce the time required for mechanical reviews conducted by the Elite Anti-Collusion Working Group on datasets stored in the public procurement platforms SECOP I and SECOP II.

This tool has been generating alerts regarding potential anticompetitive practices in government contracting since December 2023. It facilitates work by generating alerts and issuing three different types of alerts to better classify the data it produces.

**2. SABUESO**

A data analytics tool designed to detect price behaviors in products sold by large retailers and airline tickets offered by market participants in passenger aviation.

The Sabueso-Large Retailers data analytics tool allows monitoring of various products offered by different retailers (also known as chain supermarkets) and generates alerts of potential anticompetitive practices. Previously, it monitored one chain supermarket, and it is currently monitoring five.

Additionally, the new Sabueso-Flights tool enables monitoring of airline tickets offered by airlines and metasearch engines, generating possible alerts of potential anticompetitive practices. In 2023, it analyzed the behavior of over 9 million airline ticket data across 76 domestic routes, generating 180 alerts.

**3. INSPECTOR**

A data analytics and artificial intelligence tool that generates alerts for the Competition Advocacy Group regarding regulatory projects that can be examined from the perspective of free economic competition.

As of December 2023, this tool monitors 69 government entities. Several improvements have been made to notifications, including: (i) the quality of the title sent within the notification of new regulatory projects, (ii) increased proportion of

notifications with titles, (iii) the notification sent to the official now includes publication and comment closing dates for each regulatory project, (iv) improved precision and accuracy of free competition impact considerations included in the notifications issued by the tool, and (v) notifications issued by the tool include the regulatory project as an attachment.

**4. BÚHO**

Its objective is to support and facilitate the identification of news related to potential violations of the free economic competition regime in various digital media outlets. The BUHO application sends daily notifications to different officials of the Delegation for the Protection of Competition with news found in La W, RCN Radio, and Las 2 Orillas (major media newspapers in Colombia). The goal is to identify news that may indicate behaviors contrary to free competition, considering that the SIC has the authority to initiate investigations both upon request and *ex officio* against market agents.

**5. SISEC**

In collaboration with the Information Technology Office (OTI), the implementation of the Monitoring and Compliance System (SISEC) was developed. The purpose of this enhancement is to ensure transparency and efficiency in monitoring activities, data consolidation, and tracking compliance with guarantees, conditions, orders, and instructions from the SIC.

The scenarios in which the Deputy Superintendence for Protection of Competition has incorporated computational competencies to facilitate the fulfillment of its functions involve **(i)** the adaptation of its daily processes and **(ii)** the creation and use of more sophisticated technological tools. These include five programs that feature advanced programming and even artificial intelligence. Each of these tools is focused on a specific task or mission of the Deputy Superintendence or a specific area within it.

## Czechia
## Office for the Protection of Competition

### I. Project Datacros II

The Office is constantly trying to improve its methods, find new and more efficient ways of detection by testing different tools, software etc., keep up with the increasingly sophisticated practices of undertakings.

We are trying to create detection methods for use in practice, we have staff who focus directly on this issue. The Office has previously participated in a project in which a program for downloading data from the Public Procurement Portal was created, and seeks to improve and explore other detection options. For example, we are currently focusing on the extension of the input database for bid rigging detection, the creation of more advanced tests for detection, based on econometric tests and machine learning algorithms. That is why we participated in the project Datacros II. We saw in it the potential to improve the detection of prohibited agreements. We have tested this project in practice and evaluate it positively.

The University Cattolica del Sacro Cuore within which this project originated has submitted an application for Datacros III to the Commission. The project should be extended to other entities. The specific objectives of DATACROS III will be to i) expand the scope of the current DATACROS data ecosystem, ii) implement AI – driven advanced analytics functions, iii) pilot the tool-box in operational use-cases, iv) establish a multi-stakeholder cooperation and training platform and v) assess the potential benefits and constraints of the use of AI tools in organized crime and financial crime.

The main goal of Datacros III will be to develop an empowered AI-driven risk assessment and investigative tool-box, which will boost the capacity of a wide variety of public authorities (including competition authorities) to investigate organized crime networks and their business models across a variety of financial domains and criminal assets.

### II. Program Nuix

The Office uses forensic program NUIX, installed on its own devices (laptops), to search mailboxes during dawn raids. It has had positive experiences with this tool.

The investigation itself is carried out as follows. The undertaking's mailboxes are imported into this program and searched. However, the search method is always chosen to ensure the highest possible efficiency and speed of the search. Therefore, if it is more efficient to search the undertaking's mailboxes directly in his device, the search is carried out in the device.

## III. Strategy for the digitization of public procurement (Ministry of Regional Development)

This strategy, which is underway at the Ministry of Regional Development, is based on the need of the Czech Republic to address the information support of the public investment process in a systematic way and in connection with the development of eGovernment in the Czech Republic and the EU. For the period 2022 to 2030, the main mission of the development of the digitization of public procurement is primarily:

- simple and efficient public procurement and bidding,
- maximum transparency of the public procurement market and openness of systems,
- promoting innovative and correct e-procurement.

In order to maximize the circulation of data, the historical IS (Public Procurement Portal, Public Procurement Information System) have been rebuilt. The modification was related to machine-readable data, changes in legislation, etc. The Ministry of Regional Development also plans to create a data link to one database where the Office will have access.

## IV. Bid rigging e-learning project

Bid rigging is one of the most serious competition law violations, and it is generally estimated that a bid rigging cartel will increase the price of the goods (services or works) sought by 10 to 50%. Over ten percent of the gross domestic product of the Czech Republic (in the order of hundreds of billions of crowns) is invested in public procurement every year. Equally problematic is the fact that bid rigging is often associated with other offences (especially corruption), is usually well and long-term organized, and may appear to be part of fair competition. The role of contracting authorities is crucial for detecting possible bid rigging agreements, as they can both (i) prevent bid rigging agreements and (ii) detect bid rigging

agreements and subsequently inform the Office of their suspicions. The Office has therefore developed an e-learning tool on bid rigging and offers it as part of its training to contracting authorities.

Contracting authorities will learn in particular:

- why bid rigging is so harmful;
- what are the manifestations of bid rigging;
- how bid rigging can be detected;
- bid rigging in the practice of the Office (best practices);
- why the role of contracting authorities is crucial—what to do if bid rigging is suspected;
- the possibilities for private enforcement of damages.

The project was developed in the online tool Canva, using Google forms to process the tests, and therefore required very low costs, given the Office's budget.

# Denmark
## Danish Competition and Consumer Authority

## I. Introduction

Approximately five years ago, the DCCA initiated a strategic move by recruiting its first data scientists. The intention behind this recruitment was to develop new capabilities for detecting and investigating antitrust behavior. The decision to employ this new profession in the authority has indeed proven successful. Over the years, it has become increasingly evident that integrating computational skills into our workflows can significantly elevate both the quality of our work and operational efficiency across a diverse spectrum of tasks.

In this contribution, we will describe some of the notable achievements we have accomplished thus far. Part II will focus on Computational Ex Officio Analysis, and part III addresses Computational Analysis of Case Data. Additionally, part IV gives insights into potential future developments that lie ahead in this dynamic field.

## II. Computational Ex Officio Analysis

### Bid Viewer

Competition authorities worldwide use various methods to detect collusion, including whistleblower reports, anonymous tips, and *ex officio* analysis. In addition to these traditional approaches, we have developed a powerful computational screening tool called "Bid Viewer." This tool leverages cutting-edge techniques, including machine learning and artificial neural networks, to scrutinize large public procurement datasets for potential collusion.

Bid Viewer systematically analyzes bidding data to uncover suspicious bidding patterns. If competitors coordinate, certain systematic behaviors emerge, which can be detected through computational methods ("screens"). These screens flag tenders and companies exhibiting potentially collusive bidding practices.

Flagged tenders and firms are then subject to further investigation by agency staff using traditional tools and techniques. Computational screens serve as an initial step, informing decisions on whether to initiate a formal case process.

Bid Viewer relies on historic and current public procurement data from multiple authorities. By combining empirical evidence of cartel behavior with theoretical and empirical models, we enhance our ability to identify collusive activities.

Bid Viewer represents an advancement in antitrust enforcement, bridging legal expertise with computational power. Its application across jurisdictions contributes to a global effort in combating collusion and promoting fair competition.

## III. Computational Analysis of Case Data

ImageClassifier

The ImageClassifier is a meticulously trained ResNet50 model that analyzes approximately 420,000 images, encompassing both documents and non-documents. These images include screenshots of conversations, whiteboard photos, scanned documents, handwritten notes, and various other visual content such as people, pets, food, and buildings.

The primary purpose of using an image classifier, rather than relying solely on optical character recognition (OCR), is to mitigate misreadings of text that may be challenging to interpret accurately using OCR algorithms. Leveraging the ImageClassifier offers several advantages, particularly in terms of throughput. Another use case is to identify OCR candidates in a large dataset with many ordinary photos.

The model is deployed on a local Linux server within the same air gapped network as the servers running NUIX (our investigation software). Images are sent to the service, and the classifier returns a document probability score as a response. This document score integrates into NUIX as custom metadata, enhancing our investigative capabilities.

The ImageClassifier plays a crucial role in accurately and efficiently classifying visual content encountered during antitrust investigations.

Synonym Expander

In our investigative work, we utilize the Word2VecTrainer in conjunction with the SynonymExpanderService to enhance searches through extensive textual data collected during dawn raids. The primary objective is to assist investigators in

locating relevant evidence, especially when encountering specialized terminology or unfamiliar language used by the individuals under investigation.

Companies often employ esoteric vocabulary for various reasons. This may stem from domain-specific knowledge inherent to their occupational field or deliberate use of ambiguous or coded language to conceal potentially questionable content.

We vectorize words from the corpus of texts using either Word2Vec or FastText. Each word receives a vector that captures its semantic meaning. These word vectors enable various operations, including comparison between vectors, calculation of distances, clustering, and serving as input for machine learning models.

Power of Word Vectors - Consider the following illustrative calculations:

$$\text{King} - \text{Man} + \text{Woman} \approx \text{Queen}$$
$$\text{Paris} - \text{France} + \text{Japan} \approx \text{Tokyo}$$

These demonstrate the ability to capture semantic relationships using word vectors.

We train a shallow neural network to calculate word vectors. The input consists of context windows around a word (typically plus/minus 5 words). Ideally, the word in the middle of the context window serves as the output. The hidden layer in the neural net captures the abstraction needed for semantic content representation, and the extracted hidden layer vectors are used as word vectors.

Within NUIX Workstation, users input a word into the search field. By clicking a button, a Ruby script executes. The script communicates with the SynonymExpanderService, which operates as a FastAPI service within a Docker container. The SynonymExpanderService utilizes one or more models trained by the Word2VecTrainer. It identifies words used in similar contexts to the user-inputted word. Notably, the model outputs are not strict synonyms in the dictionary sense but rather words associated with comparable contexts.

Our approach enhances search capabilities by leveraging word vectors and the SynonymExpander, enabling investigators to navigate complex language and uncover relevant evidence effectively.

OSINTipede

The OSINTipede project serves as a valuable tool for conducting open-source investigations related to businesses. Through its minimalist interface, users can input various resources, including VAT numbers, article links, search queries, and employee data. These diverse inputs are seamlessly integrated into a unified Neo4j graph, which users can explore with minimal graph querying expertise, thanks to a selection of predefined query templates.

The central feature of the OSINTipede project is its graph structure, which provides several benefits:

- **Efficient Access to Information**: Investigators gain easy access to a wealth of intelligence related to a specific case.
- **Holistic View**: The graph offers a comprehensive overview of relevant attributes and relations.
- **Company Relationships**: Users can explore connections such as CEOs, board members, accountants, and other key personnel.
- **Historical Context**: Historical names, addresses, and changes in ownership are readily accessible.
- **Entity Co-Occurrences**: By analyzing articles, we can identify instances where entities (e.g., companies or individuals) have been mentioned together in newspapers or similar media.
- **Web Search Integration**: Quick search queries can reveal whether specific terms appear in Google or Bing results alongside any entities in the graph.
-

The OSINTipede project empowers investigators by consolidating diverse data sources into a coherent graph, enabling efficient exploration and extraction of critical insights for business-related inquiries.

## IV. Future Plans

Chatting with data from dawn raids

In our ongoing research, we delve into the possibilities within the realm of RAG (Relational Analysis of Graphs) applied to communication data obtained during dawn raids. The nature of communication data introduces complexities, particularly in the case of emails, which often contain implicit references, a mix of languages, and divergent communication threads. Additionally, extensive preprocessing is

necessary to structure the data appropriately for RAG. This preprocessing involves tasks such as date formatting, removal of metadata from email bodies, and elimination of auto-generated signatures.

Our ultimate goal is to adopt a GraphRAG approach. In this framework, a graph is constructed with entities (such as individuals and companies) and the objects they reference (documents, meetings, places) as nodes. These computational techniques enable us to uncover implicit context within raw communication data, enhancing our antitrust investigations.

Investigations Platform

Ultimately, we wish to include a vast majority of tools in a user-friendly platform for investigations. The idea is to host this on a local server, exposing all the tools in a single platform accessible in a browser from the servers on which our investigators already work in NUIX. It should be possible to use the SynonymExpander chat with the data using RAG, visualize subgraphs based on the data, and more. Our vision is to have a supplement to our primary investigation software, NUIX, that can handle all of the more advanced graph and NLP needs.

## V. Conclusion

Over the past five years, integrating data science into our antitrust investigations has greatly enhanced our capabilities. Tools like Bid Viewer, ImageClassifier, Synonym Expander, and OSINTipede have improved our detection and analysis of antitrust behavior. These tools use advanced computational techniques to process large datasets efficiently and accurately.

Looking ahead, we plan to refine these tools further and develop new capabilities, such as the GraphRAG approach for communication data analysis. Our goal is to integrate these tools into a unified platform, providing a seamless and advanced investigative interface.

Our computational antitrust initiatives have been successful, demonstrating the critical role of data science in modern antitrust enforcement and our commitment to promoting fair competition.

## Finland
## Finnish Competition and Consumer Authority

The Finnish Competition and Consumer Authority (FCCA) is committed to improving its effectiveness through the adoption of sophisticated computational tools. In particular, the authority is exploring methods to employ artificial intelligence, including Large Language Models (LLMs), in its daily operations. The authority is providing its staff with the tools and training to use LLMs effectively. In addition, it deploys computational tools in areas such as cartel detection and the maintenance of competitive neutrality.

A primary objective of the FCCA's competition division is the identification and elimination of cartels, given their significant and detrimental impact on competition. To this end, the FCCA has instituted a two-stage screening process to detect cartels from bidding data, with a specific emphasis on public procurement tenders. This process leverages automated computational tools to optimize efficiency and precision.

The initial stage of the cartel screening process involves the systematic application of various methodologies to a large dataset on public sector tenders, with the aim of identifying potential collusion indicators warranting further scrutiny. This stage is largely automated, employing a wide array of generalizable tests with varying thresholds. As a result of this first stage, certain markets are flagged as potentially suspicious.

The subsequent stage entails a more comprehensive analysis of the markets initially flagged. Since the markets in the database are very heterogeneous, this step of the analysis requires a higher degree of customization and a reduced reliance on automation. The objective is to distinguish between genuinely suspicious markets and those flagged due to false positives. The guiding principle is to ascertain whether the results from the initial stage could be attributed to factors other than collusion.

Upon completion of the second stage, any firms or markets still deemed suspicious are internally reported to the authority's cartel detection function, which determines the subsequent steps of the investigation. The quantitative analysis throughout the screening process is undertaken by the authority's economists, ensuring a robust theoretical and practical foundation.

The FCCA is also responsible for the oversight of competitive neutrality regulation, aimed at ensuring fair competition between the public and private sectors. The principle of competitive neutrality means that public businesses should not have any undue advantages over their private competitors because they are government-owned. The authority is frequently consulted by public sector entities seeking to ensure that they are not infringing the regulations.

In order to be able to respond more efficiently to these queries, the FCCA has piloted a chatbot trained on a dataset comprising previously asked questions and their responses, as well as question-answer data from the authority's web pages. The diversity of the questions necessitates a more complex approach than simple few-shot inference. Yet, the training data is not sufficiently extensive to allow for fine-tuning the model. The approach is currently being tested in a 128k token GPT-4 LLM, which can process all training data and assist in drafting an expert response. Future development plans include a combination model based on Azure Prompt Flow, capable of handling larger datasets with embedding and web search capabilities.

The objective is not to provide advice directly to the inquirer, but to expedite the creation of a response by the authority's staff. The model operates within the Finnish government's private cloud, with no data directly used to train the LLM. While currently applied only to competitive neutrality data, the success of the pilot program could see this setup extended to other customer inquiry response needs.

In addition to the above projects, the FCCA is also leveraging web scraping for a broad spectrum of purposes related to both antitrust cases and policy reports. Web scraping allows for the extraction of large amounts of data from websites quickly and efficiently, and is typically used in cases where alternative data sources are unavailable.

## France
## Competition Authority

In 2020, the Autorité decided to create a dedicated digital unit in order to strengthen its resources in digital areas. Launched in September of the same year, the digital economy unit is composed of four people (including two data scientists) and among its objectives, one of them is to develop new digital and computational tools allowing the Autorité to deepen its understanding of the digital sectors and to facilitate the work of the cases handlers.

The Autorité is proud to be part of the Stanford Computational Antitrust Project and to have contributed to its first two annual reports. The following presents what has been done since then.

As a reminder, the Autorité, in collaboration with CodeX Computational Antitrust, has published on its website the first interactive network graph tool capable of identifying within the Autorité's publications (meaning the decisions, opinions, and interim measures published between 2009 and 2021) the references to its other publications and to represent these interconnections in a graph.

An article explains all the process and the visualization tool developed by the Autorité along with the entirety of the data are accessible to all as open data on the Autorité's GitHub.

This tool has been replicated for the 3,275 merger decisions published by the French Competition Authority up to 1st January 2024. Its code is now available on the Autorité's Github.

A high-resolution image of the corresponding graph is also available here until the interactive tool is made available online.

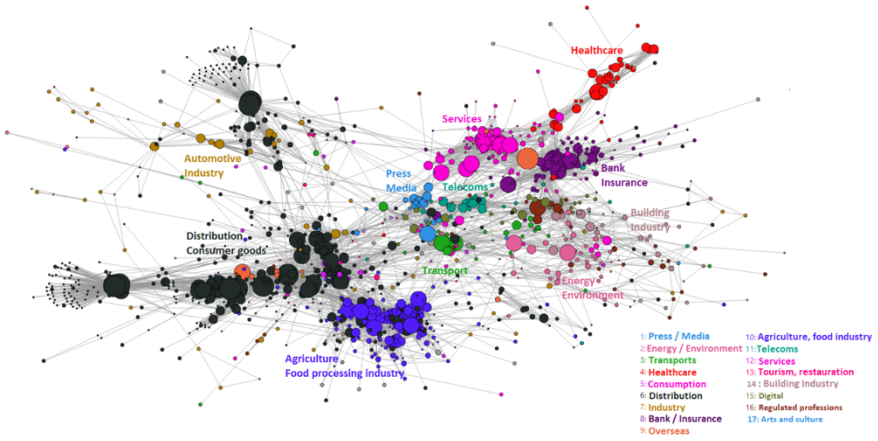


**Figure I:** Graph of the Autorité's merger decisions

The Autorité has also built an in-house search engine for the use of the Autorité's staff and has enriched it with the previous knowledge gained from its work described above on detecting references. The detecting tool has also been improved as it no longer only detects references to the other publications of the Autorité, but also to those of additional entities useful for our case handlers (like publications from the Commission, from French or European courts, etc.)

One example is shown below where one can see that regarding the decision 21-D-17, it quotes one decision from the Paris Court of Appeal 8 times (and the link to the decision is also provided).


**Figure II:** Presentation of Decision 21-D-17 in the Autorité's in-house search engine

The Autorité has also adopted an open-data approach. It has made available all its publications (in French), issued since 15 March 1988, i.e. more than 5,700 documents, available in a single database.

Under the Etalab 2.0 Open Licence, it is possible for anyone to re-use and appropriate it (for example, to power a search engine or to fine-tune a Large Language Model).

For our English readers, all the decisions, opinions, and press releases translated into English by the Autorité are also available in a single database, accessible through the QR code provided below.

**Figure III:** QR code to access the databases

Finally, the digital economy unit is currently developing a variety of automated tools for ongoing cases investigated by the competition units within the Autorité. We will share the technology we have developed for those cases in due course.

## Netherlands
## Authority for Consumers and Markets

The Netherlands Authority for Consumers and Markets (ACM) is charged with competition oversight, sector-specific regulation of several sectors, and the enforcement of consumer protection laws. Data and technology-related work and development within the Dutch Authority for Consumers and Markets (ACM) is mainly centered within the Taskforce Data and Algorithms (TDA). This department focuses on the development of analytical tools and analysis to support the several goals the ACM holds as an organization. The focus of this article will be on the work the TDA does for the competition department within the ACM, as this department has a large focus on antitrust.

Mainly the TDA developed and is still improving two tools that are being used or will be used by the competition department: the *Tender-Tool* which helps identify bid-rigging behaviour, and the *Signalling Monitor* which detects price signaling in news. Besides these tools, the TDA supports frequently data analysis for merger and acquisition requests, and cartel research.

Furthermore, we are currently researching the implementation of large language models, vision models, and more generally, foundation models into our detection and case-research work.

### 1. Tools

*The Tender Tool* has been developed over the recent years and is maintained primarily by the TDA. The tool was designed to facilitate co-development and ease of adoption by providing the functionality in a package. The Authority for Consumers and Markets competition department is the primary internal user of the Tender Tool, and the main usage is to identify aberrant bidding behavior.

The Tender Tool is shared within the ECN and from there can be used by other European Competition Authorities. There is a keen interest in handling the problem of "bid rigging" in a more coherent and unified manner from the European Commission, which the Authority for Consumers and Markets is supporting through collaboration on the continued development of the Tender Tool.

*The signalling monitor* has recently been developed and implemented to detect signals of price changes in news articles. The signaling monitor uses the BERT Language Model, a deep learning model using complex natural language processing.

### 2. Data analysis toolkits

The TDA participates frequently in antitrust investigations. These analyses differ in three main directions:

- **Assisting in merger and acquisition requests**: providing analysis of geographical markets.
- **Analysing pricing behavior** to investigate price arrangements or to measure the effects of an intervention.
- **Analysis of vertical price arrangements**.

### 3. Data analysis frameworks

Besides these innovations, the ACM and particularly the TDA have been adopting several frameworks like privacy assessments and Impact Assessments for Algorithms (The Netherlands IAMA framework), and will keep doing so as new directives develop, for example the AI Act. These frameworks will help the department and the ACM even more to make informed and ethical decisions about the use of data and algorithms in general.

## Poland
## Office of Competition and Consumer Protection

As part of an EU-funded initiative, the Polish Competition Authority (UOKiK) developed ARBUZ, an AI-powered system based on supervised machine learning, between 2020 and 2022. This system was integrated into the organization's daily operations in early 2023.

ARBUZ operates on a core principle: it assesses the semantic similarity between segments of a contract under review and clauses that have been previously categorized by UOKiK as either abusive or non-abusive (safe). The system was trained using a structured database meticulously compiled by UOKiK employees, which included annotated excerpts from court rulings, decisions by the President of UOKiK, and 'soft calls'—informal requests made to entrepreneurs to cease using prohibited contract terms voluntarily.

The system analyzes standard contract terms, identifying potentially abusive sections by referencing verified clauses from the database. As it relies on supervised machine learning, the analysis of contract terms also serves to automatically update the original database.

In 2024, UOKiK released a 'White Paper Advancing Consumer Law Enforcement with Artificial Intelligence.' This document discusses the organizational, legal, and ethical aspects of creating and implementing AI-based systems within public consumer protection agencies. It also explores potential future applications and directions for the development of such technologies.

Dark Patterns

UOKiK has launched an EU-funded project 'Detecting and combating dark patterns with Artificial Intelligence' to develop a methodology for conducting proceedings on dark patterns and to explore opportunities for deploying AI in consumer protection. This includes the development of a Proof of Concept for an AI tool capable of detecting dark patterns, as well as research into the latest trends in AI and machine learning.

As part of this project, in 2023, websites of online stores and other service providers were screened, and sectors with the highest occurrence of dark patterns were identified for further analysis.

Preliminary proceedings will be initiated based on findings from open-source intelligence, which will include neuromarketing tests designed to study neurobiological human reactions to dark patterns.

During the course of administrative proceedings, UOKiK plans to organize an IT analysis of websites conducted by external experts and to conduct consumer surveys about online purchasing experiences and the influence of specific dark patterns.

# Saudi Arabia
# General Authority for Competition

## I. Introduction

Computational antitrust can assist antitrust agencies in navigating increasingly complex and dynamic markets. Implementing computational methods can enhance antitrust agencies' ability to detect, analyze, and remedy anticompetitive practices. These tools can also simplify merger control and related procedures.

In October 2017, the Kingdom of Saudi Arabia's Council of Ministers adopted Resolution No. (55) to change of the name from the "Council of Competition" to the "General Authority for Competition" (the "Authority") and to approve the Authority's Statute. The Authority continuously employs and improves various tools to enhance market monitoring and enforce competition laws.

## II. Business Intelligence (BI) platform

The Authority has implemented a Business Intelligence (BI) platform to visualize its market analysis. The platform serves as a central repository for market indices and insights, encompassing a wide range of industries, in collaboration with public institutions and private data providers. The BI platform provides insightful assessments of sectoral competitiveness. Key metrics such as the number of market players, their market size, revenue figures, and other relevant data points are readily available through the platform. This comprehensive overview empowers the Authority to gain an in-depth understanding of the competitive landscape within each sector.

While the BI platform offers significant advantages, it also presents certain challenges. A critical consideration lies in integrating data from diverse sources. Data inconsistencies and varying formats can pose challenges to seamless integration. Furthermore, establishing data-sharing agreements with private providers requires careful consideration of data privacy and security concerns.

In addition to the sectorial assessments, the platform visualizes the analysis of specific areas of interest to competition, such as public procurement, goods prices, and consumer sentiments.

## III. Bid-rigging Screening

Currently, the Authority uses two levels of screening. The first screening focuses on pricing. It employs mathematical and financial relationships to screen market price patterns for potential indications of collusion. Supervised machine learning models, trained with existing screening data, are used to predict the risk of collusion by recognizing patterns that are not immediately apparent through traditional analysis methods. This allows the Authority to efficiently and accurately screen large volumes of data, improving the probability of identifying cartels at an early stage. The results of the analysis are streamlined on the BI platform and shared internally with other enforcement departments with the Authority.

The second screening uses market-level analysis to detect behaviors such as market sharing, rotation, or other forms of coordination, considering the market structure of the industry. Advanced econometric models help identify or justify pricing or quantity anomalies, demand growth, and other irregularities indicative of collusive behavior before reporting them. The Authority continuously refines these methods and improves them by collaborating on advanced economic and statistical research. This integration of the latest innovations from economic studies enhances detection techniques, keeping the Authority at the forefront of cartel detection.

Additionally, the Authority is set to integrate the use of advanced AI with its existing tools to improve cartel detection. One important method is network analysis, which will help the Authority to understand the relationships between market participants in different markets (such as the public procurement market). By analyzing publicly and commercially available data, such as commercial registrations within different tenders, it can draw connections to identify complex patterns and relationships that may indicate collusion. This analysis allows the Authority to visualize connections between different market participants, identify unusual clusters of activity, and detect anomalies that could signal cartel formation or existence, and thus possible infringements.

If a matter is under investigation, and documents are obtained during dawn raids, the Authority will incorporate this new data with pre-existing data to map the relationships between assumed cartel members. This method ensures that any suspicious patterns are identified and backed by robust data.

Moreover, the Authority uses its own classification system of industries on the basis of industry concentration and is set to use a less aggregated product-level

supplier concentration classification system for most of the products and services frequently referred to in public tender procedures. This classification will help not only to connect the lines between participants in different markets under investigation but also to set an *ex ante* procedure to identify markets that have the greatest risk of collusion based on some indicators for collusion identified in literature. This will enable appropriate advocacy policy to reduce these risks.

## IV. Prices

The price monitoring platform is a data-driven approach to tracking and analyzing pricing data across a range of consumer products locally and within the GCC region. It uses various price volatility and brand concentration indices to reveal important information related to market power, pricing coordination, demand growth, and market trends. The platform's regional scope provides valuable insights into pricing dynamics across the GCC region, enabling a more holistic understanding of market competition through price and brand benchmark indices.

## V. Consumer sentiment

The Authority employs a public opinion monitoring tool that tracks and analyzes public sentiment on competition issues across various media channels, including social media, television news, and newspapers. The tool uses keywords to identify content relevant to the Authority and the enforcement of competition law. Currently, around 55 words are incorporated, and this pool is frequently updated. Examples of keywords including tying, bundling, abuse of a dominant position, and monopoly. This tool allows the Authority to identify emerging competition issues that cause public concern. Such insights are considered when prioritizing investigations and enforcement actions, ensuring that the Authority focuses its resources on areas of greatest concern to consumers.

## VI. E-platform

The Authority uses a dedicated E-Platform "Monafsah" for managing Mergers & Acquisition (M&A) notifications and applications. This platform streamlines the entire process, enhancing efficiency and transparency. Complex applications can also be broken down into smaller, more manageable inputs for thorough analysis. Importantly, the E-Platform extends its functionalities beyond M&A, serving as a centralized system for managing settlements, leniency programs, complaints,

surveys, and exemption requests. This centralized approach fosters efficient case management and reduces administrative burdens for both the Authority and applicants. Additionally, the platform serves as a valuable repository of M&A information and statistics internally and through open data protocols, providing data-driven insights for future case reviews and informed decision-making. This data can also offer valuable insights into the Authority's internal and external processes related to M&A regulation, facilitating continuous improvement.

## Singapore
## Competition and Consumer Commission

In 2023, the Competition and Consumer Commission of Singapore ('CCCS') established a new Data and Digital ('D2') division to lead and focus the agency's efforts on technology development in big data and digital markets. Below is a list of tools recently rolled out or currently under development by D2.

### I. Complaint Analytics Tool (CAT)

The CAT is a python-based tool, developed in-house by D2, which applies natural language processing on a sizeable volume of text passages, and uses machine learning to cluster them into intelligible topics, which are then ranked by popularity. The tool can also compare topics across time and trace the evolution of a particular subject over a period.

Since its introduction, the CAT has found several use cases. First, it has been used to analyse complaints received from the public. The results are regularly presented to senior management for a deeper understanding of trending issues in the marketplace. This also helps to generate leads for potential enforcement cases. Second, the CAT has been applied to analyse feedback received in some public consultations to improve productivity in considering such feedback. Third, the tool has also been applied to analyse entries from the CCCS-Economics Society of Singapore ('ESS') essay competition to identify frequently-discussed issues.

### II. Fake Review Detection Tool (FRDT)

As CCCS is the enforcement agency for both competition and consumer protection laws, D2 has developed the FRDT to augment our latter function. The FRDT is a python-based machine learning model that underwent supervised training with open-source, labelled textual data on genuine and fake reviews. The tool has been applied to certain ongoing investigations involving digital platforms to analyse a large volume of online reviews and identify suspicious ones.

Again, this is a generic tool which learns to identify text passages with certain characteristics as required by individual use cases. The model can then be trained with corresponding data that are relevant for the model to learn about that required

characteristic. In the future, it is conceivable that the same tool can be applied to detect suspicious tenders involving bid-rigging, or even resumes in the agency's recruitment exercises to identify potential candidates for interviews.

## III. AI Verify Toolkit (AIVT)

CCCS is a member of the AI Governance Roundtable, which is a committee between government agencies to devise AI governance principles in Singapore. The AI Verify foundation is a collaboration between technology partners in the public and private sectors to promote best practices and standards for AI. The AI Verify Toolkit is an open-source AI testing tool developed by the AI Verify foundation to test the compliance of AI algorithms with Singapore's AI governance principles.

In collaboration with the Infocomm Media Development Authority of Singapore (IMDA), the Personal Data Protection Commission (PDPC), and the AI Verify Foundation, CCCS is currently developing a customised "vertical" of the AI Verify Toolkit which tests the compliance of an AI algorithm with competition and consumer protection principles in Singapore. It is envisaged that businesses can incorporate the vertical into their compliance programmes, on a voluntary and self-assessment basis.

## Slovakia
## Antimonopoly Office

### I. Economic Analysis Data

With growing big data, accessibility The Antimonopoly Office of the Slovak Republic (AMO SR) started to use STATA 18.4 and R/Rstudio for more advanced data analysis. Stata is a powerful statistical software that enables users to manage large datasets, produce statistical analysis and graphical visualizations of data. It is primarily used by researchers in the fields of economics, biomedicine, and political science to examine data patterns.

In addition to STATA, analysts at the AMO SR also plan to work with R/RStudio which is used in data analysis to import, access, transform, explore, plot, and model data, and for machine learning to make predictions on data.

At the AMO SR we plan to primarily manage and analyse large economic data such as business register micro firm-level data, large own-firm accounting datasets, public procurement data, state aid data at the level of the recipient (businesses) and individual-level labor market data to provide evidence-based support in investigations carried out by the AMOSR. We also plan to publish regular economic briefs and materials on the state of competition in the various sectors of economic activity.

### II. Forensic Tools

During the inspection, the Office obtains evidence of illegal anti-competitive conduct by the undertakings using specific IT technology with elements of forensic analysis. Digital forensic activity preserves the credibility of obtained digital traces and evidence, which are often subject to a court review.

The use of special tools and procedures in the identification, acquisition/restoration, processing, and analysis of data/information stored in electronic form by the undertaking guarantees that the obtained evidence is clearly derived from the obtained digital information (at the time of acquisition) and has not been altered in any way (integrity). The Office carries out inspections, or forensic activity at the premises of undertaking. The choice of software and hardware by the

Office is also adapted to this method of data acquisition and processing and to financial possibilities, with the aim of processing as much data as possible within the set time.

Due to the growing complexity and advanced level of IT systems, which undertakings use not only to store data, but also to conceal illegal evidence, the Office is aware of the need to use adequate tools for forensic data analysis, taking into account the effective use of funds for their procurement.

Currently, the Office has 6 pieces of mobile stations, which are powerful enough to work with complex forensic analysis tools, including other accessories (mobile printers, writeblockers, card readers, reductions, external disks and others).

Forensic analysis of mobile phones is performed using the UFED device with the appropriate software equipment (Cellebrite Physical Analyzer).

The main software tools of the Office are: FTK Imager (a tool for automatic creation of guaranteed copies of data - image), OS Forensics (a complex analytical tool designed for forensic data processing), and KillDisk (a tool for guaranteed deletion of data - wipe).

The Office also emphasizes the continuous training of employees ensuring the Office's forensic activities.

In an effort to adapt to current trends in the field of communication and data processing, the Office continuously reviews the state of forensic equipment. Based on the analysis, the Office concluded to improve/supplement the Office's technical equipment with a more robust analytical tool—NUIX, which is also used by other European competition authorities and shows high efficiency in processing data of various formats and sources.

## III. Website Traffic Monitoring

SimilarWeb is an online platform that provides insights into website traffic, competitive analysis, and digital marketing. The platform collects and evaluates information from a multitude of diverse sources, which it then publishes on its website in a processed form. Collaborating platforms that provide data for SimilarWeb include portals such as Google and Facebook. SimilarWeb states on its

website that it has information on more than 100 million pages from over 190 countries.

For the Office, the most crucial function of SimilarWeb is the website traffic monitoring service, which tracks page visitation. In this category of services, SimilarWeb offers information about the number of visits to a specific page, unique visitors, the duration of page views, and even details on whether a visitor viewed the page from a mobile or PC browser. This information is not freely available from any other source.

Digital markets pose a challenge for the office in accurately determining market shares, which is essential for the proper assessment of the situation in a specific digital market. In the past, as part of its investigation, the Office used SimilarWeb to compare the traffic and popularity of different websites with the same or similar focus in order to determine their market share in the digital market. A tool such as the SimilarWeb platform could be a means to improve the Office's analytical capabilities.

## IV. Public Procurement Analysis Data

In October/November 2023, the Slovak Competition Authority had access to the Transparex information system (database) available for testing—TRANSPAREX—Najväčšia databáza ekonomických a obchodných informácií. Transparex is probably the largest database of Slovak and Czech companies which mainly provides information on all legal entities, proprietary and personnel connections and conflicts of interest, public procurement data, a database of new orders, references from contract clients, information from the Land Register, data of the Register of beneficial owners.

The biggest advantage of this database was that we could access comprehensive information in one place about proprietary and personal connections of companies, their economic and financial indicators, but also important warnings about risk events, their participation in public tenders or their success. This was very helpful in identifying possible collusion in the public procurement process. The information was clearly processed and working with it was not time-consuming.

The information in this database is publicly and freely accessible but not in one place, so searching for information through this database saves time which was probably the biggest advantage for us.

The Transparex database would have been very beneficial to our work but due to the price and our financial capabilities it was not possible to purchase it.

## V. Artificial Intelligence

Given the particularity of competition law and competition policy, it would be appropriate to have a suitable AI-based tool that would enable case-handlers to effectively search, link, and apply data from their own database of competition decisions, soft law, and jurisprudence.

The Office therefore has focused on the possibility of introducing an AI tool in the form of a user-friendly chat to search for and evaluate the application practice on two levels. The basic level is the database of non-public (full-format) decisions of the Antimonopoly Office usable for AI goals and an additional level represents the database of public decisions and jurisprudence of the European Commission and the Court of Justice of the EU with the possibility of expansion using other databases of important NCA ' s decisions.

During the first half of 2024, the office has been looking for a suitable supplier of such an IT tool.

# Spain
# National Commission for Markets and Competition

## I. Short summary

Last year, the Spanish National Markets and Competition Commission's (CNMC) contribution to the 2nd Annual Report Stanford Computational Antitrust addressed the topic of Computational tools for detecting collusion.

Taking into consideration that the CNMC has among its top priorities adapting its competition enforcement policy to the current wave of digitisation and use of artificial intelligence, the initiatives already in course at that moment have experienced an evolution and, in some respects, a relevant breakthrough.

As it was then reported, the CNMC has an Economic Intelligence Unit (EIU) devoted to, among other activities, applying new techniques of data analysis, using artificial intelligence to detect not only new forms of (namely, algorithmic) anticompetitive conduct, but also traditional ones, subtler and more refined every day. The unit also aims to increase *ex officio* detection and strengthen the robustness of other investigations.

The availability of each day a more ample database covering all aspects of public procurement, including detailed factual data of enterprises and related persons thanks to the agreement signed between CNMC and Real Estate Registers Official College of Spain, has been also an important branch of activity for the EIU, providing our data analysis tools a more solid base to yield relevant results.

The CNMC thus remains committed to progress in applying new techniques to ensure that competition is safeguarded in all markets, especially in public procurement. The scope of this contribution will be to summarily describe the actual use of some of them and their stage of development.

## II. Detection and analysis of anticompetitive conduct in the Economic Intelligence Unit

Algorithms and other information technology tools can be used for anticompetitive purposes, by means of software tools with a varied degree of

sophistication (pricing algorithms, etc.). But anticompetitive conduct can also appear in other forms, such as self-preferencing or discrimination (e.g. in rankings or matching algorithms). These risks appear more often in digital business models where algorithmic tools are used frequently, but they can also take place in other sectors.

To tackle these challenges, the Spanish National Markets and Competition Commission (CNMC) introduced digitisation as a priority in its Strategic Plan,[8] leading to different initiatives in its annual Action Plans.[9] One of such initiatives was the creation of an Economic Intelligence Unit (EIU), to take advantage of new digital techniques as an aid to detect different kinds of anticompetitive conduct, including when this conduct is facilitated by algorithms and other advanced tools.

This contribution summarizes and updates the work carried out by the Economic Intelligence Unit (EIU) to increase the detection of anticompetitive conduct through digital tools.

The analysis carried out by the EIU is used both for purely *ex officio* detection and for refining and completing the analysis in other cases: enhancing and reinforcing evidence, adding information to existing complaints, dawn-raid planification through OSINT information, etc.

Collecting the right data allows to use specific techniques and analyses, such as supervised machine-learning, graph analysis, and other advanced techniques related to data management and processing.

Prior to the creation of the EIU itself, the CNMC worked since 2015 on the construction of a public procurement database, developing specific algorithms for data extraction and cleaning, categorizing the data by quality levels regarding all participating bids and bidders.

Regarding public procurement, the CNMC database is the result of the aggregation of different sources. Firstly, the Public Sector Procurement Platform (PSPP) is a completely up-to-date electronic platform publishing all the calls for tenders and their outcomes. It encloses data from regional and local contracting bodies, since platforms of the different government authorities and public entities

[8] CNMC, *Plan Estrategico*, https://www.cnmc.es/sobre-la-cnmc/plan-estrategico (last accessed June 8, 2024).

[9] CNMC, *Plan de Actuacion*, https://www.cnmc.es/sobre-la-cnmc/plan-de-actuacion (last accessed June 8, 2024).

are interconnected to establish a single platform that centralises the publication of public sector procurement. Thereafter this repository is completed with data from the procurement bodies that are not integrated within PPSP, such as the centralized contracts and with the non-standardized documents associated with the tenders, such as notice of award or tender amendments.

In addition to it, the CNMC also downloads the Ministry of Finance Centralised Procurement Portal information of framework agreements, centralised contracts, and other centralised procedures for minor contracts. So far, the database contains more than 3.5 million contracts and near 4 million lots.

Therefore, this CNMC's public procurement database could be considered one of the most complete at national level, with structured data about tenders from all public administration levels and types, but also the related documents (non-structured data).

In addition to all these, in 2023 there arose another relevant source of data available due to the agreement signed between CNMC and Real Estate Registers Official College of Spain. Thanks to it, CNMC has access to all the official economic and financial information about the companies acting as bidders in public tenders, such as the location of business premises, main stockholders and/or controllers of the company, successor firms in case of extinction, etc.

With the combination of data coming from all these sources, the CNMC has hence created a complete digital database of public procurement in Spain that has allowed the development in recent times of several powerful tools that take advantage of the processed data and currently help EIU to detect potential cases of bid-rigging.

In the first place, the CNMC has a comprehensive set of analysis tools supporting detection. These are mainly business intelligence tools, like Microsoft Power BI for visualization and analysis, along with other in-house developed, dedicated web-based searching tools. These are in some aspects conventional tools, but tuned to deal efficiently with great amounts of data, giving our analysts the ability for discovering subtle trends in such data, or to relate evidence from a particular tender to other activities of the subjects under investigation.

In addition to it, we have taken advantage of graph-oriented databases as a source of screening variables, based on graph metrics calculation, to serve as inputs for the

machine learning process (*see* below). Such databases store nodes and relationships, instead of tables and documents. The screens thus generated represent the degree of relationship (proximity, in graph theory terms) among activity sectors, companies, and the company leaders themselves.

Along with all of that and, more recently, access to non-awarding data (losing bids) opened the way to the use of techniques that made it possible to decisively promote pure *ex officio* detection. Eventually it led to the development and practical creation of BRAVA (Bid Rigging Algorithm for Vigilance in Antitrust).

BRAVA is a tool based on supervised machine learning that, through different models, can classify the different offers submitted to a tender as most probably collusive or competitive. The success rates obtained in the testing of these models are largely satisfactory, never less than 90%. It employs supervised learning based on raw variables (coming from the public procurement and registrars' databases already mentioned) and more than 20 screens coded, some based on international papers and academic articles and some others developed from scratch by the EIU considering its relationship with corruption / collusion.

After this process, data was divided into training and test sets, in order to proceed with the supervised machine learning process through different iterations. This process shows us the best performance variables from all the possibilities. In practice, this represents a double utility for determining collusion:

- It allows CNMC to target specific tenders, such as those included in a communication from procurement bodies or coming from whistle-blowers, and
- It enables to trigger pure *ex officio* analyses of certain sectors, especially on markets suspiciously under more stable cartels (whose members are less prone to defection).

Although BRAVA is always under continuous and intensive development, its first implementation is already in full operation, and its results are really promising.

Another investigative approach at the EIU which is quite different but takes advantage of some of the techniques mentioned above, is the use of Open-Source Intelligence (OSINT) and Human Intelligence (HUMINT) tools in order to provide an accurate identification and location of organizations and persons of interest, the relationships among them, and their degree of control of the companies and organizations which are under close scrutiny.

## III. Main conclusions

The enforcement of competition law in digital markets is complex, especially as far as algorithms and artificial intelligence are concerned. Competition policy tools are flexible enough to adapt to the disruption driven by digitization, but some challenges still remain.

The CNMC has also given strategic priority to monitoring anticompetitive conduct driven by algorithms and digital business models, thus adapting and refining its enforcement tools. There is a real need for advanced computational tools based on similar techniques, to ensure a more even 'playing field' among cartelists and competition authorities, and therefore some of the current efforts of CNMC are set on that target.

In fact, the new, AI-based BRAVA system, recently set up by the CNMC and its Economic Intelligence Unit as described above, appears to be a definite effort in such a fight against bid rigging, one of the most relevant and widespread collusive activities.