

AI: The Growing Extent of Intelligence  
Measuring Consumption Reactions to  
Political Conflict with Smartphone Location Data  
Regulatory Impact Analysis  
in Case of Unstructured Data  
Data Sharing: Don't Forget the Costs



## **Impressum**

Redaktion  
Prof. Dr. Peter Gomber  
Julian Schmidt, M.Sc.

Herausgeber  
Prof. Dr. Peter Gomber  
Stellvertretender Vorstandsvorsitzender  
efl – the Data Science Institute e. V.

Kontakt  
[info@eflab.de](mailto:info@eflab.de)  
[www.eflab.de](http://www.eflab.de)

Gestaltung  
Novensis Communication GmbH  
Weilburg

1. Ausgabe, 2023  
Auflage: 200 Stück (Print) / 2.000 Stück (E-Paper)

Copyright © by efl – the Data Science Institute e. V.

Printed in Germany  
ISSN 1866-1238

## Editorial

# AI: The Growing Extent of Intelligence

Wolfgang König

We gathered our first experiences with Artificial Intelligence (AI) 33 years ago as a small (science) kernel of a large-scale B2B application of Lufthansa with respect to optimizing the cargo by-loading on passenger airplanes (background: Lufthansa was (and still is) the world's largest air cargo carrier). Weight goods (like lead) were charged by their actual weight; voluminous goods (e.g., flowers) by a higher kilogram rate. Ideally, the airplane's cargo capacity, which is restricted by both weight and volume upper limits, is used by a ground layer of lead complemented by a thick layer of flowers, resulting in a relative increase of freight revenues of, ideally, close to the factor two. However, qualified air-freight requests do not only comprise "lead and flower" transportation and they arrive in continuous streams. The task of the AI was to properly select out of the incoming stream of airfreight applications those that substantially and sustainably raised the relative revenue over one.

We recall these "ancient" experiences because of two important take-aways: the first is that one needs a sufficiently resilient (business-) func-

tional specification, i.e., a professional concept (here: of air-freight dispatching). In German this means: "Fachkonzept" (interestingly, there is no nice word for it in English). It tells us that, independent of the underlying IT concept, under specific circumstances (in our case for instance: enough "population" in the incoming stream of airfreight applications with varying specific weights) some minimum business advancements will be realized with some probability. The application success of such an AI system is measured against the professional understanding of responsible humans.

In modern times (three decades later), massive increases in compute power and data storage capabilities open new ways of AI application. We train AI systems with historical real-world data on, e.g., multidimensional decision circumstances and success or failure of the respectively resulting decisions. They are systematized in a way to reversely calculate the (yet to unveil) inherent professional concept. This procedure is more a systematic business-functional disclosure – or even further: a scientific exploration



**Prof. Dr. Wolfgang König**  
Senior Professor  
Goethe University

process – to detect more efficient combinations of building blocks of a solution than (as in the old times) a mere extension of the human calculation and data storage capabilities.

Two examples underpin the necessity to evaluate the professional concept options, i.e., the calculations that AI systems offer (and this is the announced second take-away) also in modern times: the responsible human must understand the permanent superiority of the professional concept, not only for a handful of application cases but for a large number of different cases (which also bear some variety of input parameter combinations).

So, first, how far can we move forward by just "extending" data of the past? For instance, we see highly interesting literature on "pay attention to hidden societal discrimination in historical data", e.g., in the insurance business. In general: when business environments change quickly (as they also did in the airfreight business), we cannot only rely on what an AI system has nicely calculated based on history.

Second: today's business cases are often more complex than a lot of historical cases. Ergo: yes, the "intelligence" of AI systems increased and improves further. But also the human business intelligence increases. Apparently, both develop in a kind of lockstep – with a sustained supremacy on the human side – in particular with respect to the permanent professional concept.

A final note on what happened to our Lufthansa case: The system worked nicely as expected and was integrated in an extensive process. But our AI module was nevertheless abandoned after only eight weeks in operation: The freight dispatchers had not been properly involved in the system development, and they just refused to employ our module. Our suspicion: they used the findings via their human system interface.

More than three decades of AI progress and applications, including 20 years of efl advancements, teach us that the human (as for instance user, developer, and business manager) is the narrowest bottleneck in applied information technology.

## Research Report

# Measuring Consumption Reactions to Political Conflict with Smartphone Location Data

FIRMS, RESEARCHERS, AND POLICY MAKERS OFTEN WANT TO MEASURE CONSUMPTION AND ESPECIALLY HOW EVENTS, PROMOTIONS, OR POLICIES AFFECT IT. MEASURING CONSUMPTION REACTIONS IS OFTEN HARD. FIRMS LACK ACCESS TO COMPETITORS' SALES DATA AND REGULARLY DO NOT SHARE THEIR OWN WITH OUTSIDERS. LARGE SAMPLES OF SMARTPHONE LOCATION DATA COULD SOLVE THIS PROBLEM. THIS ARTICLE DESCRIBES A RESEARCH PROJECT USING SMARTPHONE LOCATION DATA TO ESTIMATE CONSUMPTION REACTIONS TO POLITICAL CONFLICT DURING THE TRUMP PRESIDENCY.

Celina Proffen

### Introduction

Political conflict is a fixture of modern democracies. Such conflict shapes societal and political outcomes. But does political conflict with a foreign country also influence domestic consumers' daily consumption choices? In our study, we investigate whether consumers boycott goods associated with the opposing country in a setting where the conflict does not directly influence the characteristics of these focal goods.

More concretely, we use the US-China trade conflict to analyze whether consumers reduce their visits to Chinese restaurants in the

Lukas Jürgensmeier

US when political relations deteriorate. We measure the intensity of political conflict through the negativity in media reports and rely on smartphone location data of more than eleven million devices to proxy daily visits to over 194,500 restaurants in the US.

We find that worsening US-China relations induce a statistically and economically significant decline in visits to Chinese restaurants relative to the average restaurant in the US. At the same time, visits to other foreign cuisines also decrease substantially, while visits to traditional American restaurants increase. We interpret these results as

evidence for international conflicts triggering ethnocentric consumer behavior – defined as consumers' tendency to discriminate against foreign products (e.g., Shimp and Sharma, 1987; Sharma et al., 1994).

### Smartphone Location Data as a Proxy for Consumption

We proxy consumption of conflict-associated goods using daily visits to dine-in restaurants in the US in 2018 and 2019, distinguishing among restaurants serving different types of ethnic food. The underlying data stems from Safegraph, a company that gathers smartphones' global positioning system (GPS) location data and attributes visits of these devices to specific location polygons corresponding to individual buildings. Safegraph attributes a visit to a restaurant once a device resides at the dine-in restaurant for over four minutes.

After obtaining metadata on US dine-in restaurants and information on their daily visits, we use Safegraph's sampling rate by small geographic areas to remove sampling bias and estimate the actual number of visits. We scale the raw data by monthly sampling rates for each area, which we calculate by dividing the number of smartphones tracked therein by the number of residents. We exclude all restaurants with an average daily visitor number smaller than ten since Safegraph intentionally randomizes low numbers of visits to preserve visitors' privacy.

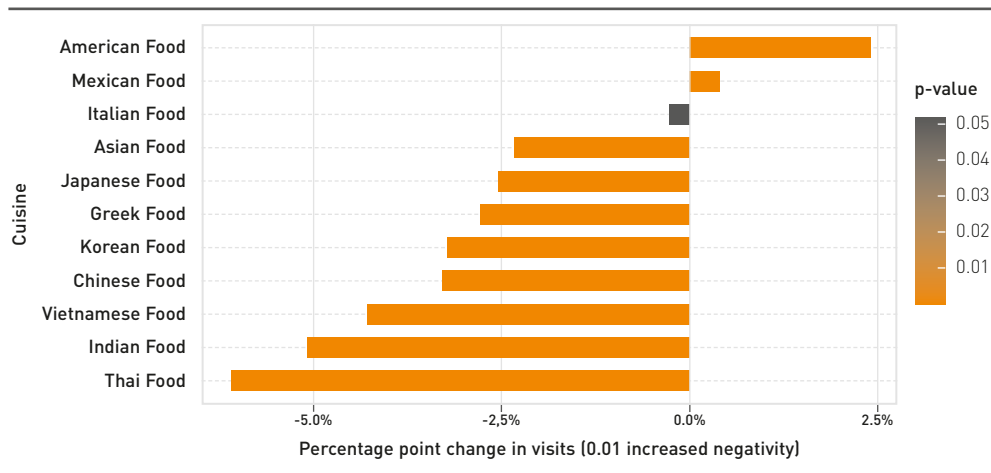
### Approach

To investigate how political conflicts influence daily consumption behavior, we exploit the context of the US-China trade conflict and measure the intensity of this conflict based on US media reports about China. More precisely, we use textual analysis methods to measure the negativity in media reporting over the past week and compute a daily metric for the intensity of political conflict with China. To measure consumption choices, we use daily data on restaurant visits in the US. Finally, we use a two-way fixed effects estimation to assess how aggravations in the political conflict with China causally affect visits to Chinese restaurants compared to the average non-Chinese restaurant.

Our baseline analysis relies on estimating the following model:

$$Visits_{it} = \gamma_i + \tau_t + \beta \times Chinese_i \times Negativity_t + \epsilon_{it}$$

Thereby,  $Visits_{it}$  denotes the number of visits to restaurant  $i$  on day  $t$ .  $\gamma_i$  is a restaurant-fixed effect that accounts for time-constant differences in the number of visits to individual restaurants (e.g., due to varying restaurant sizes, business models, or locations).  $\tau_t$  are day-fixed effects, which capture day-level variation in visits common to all restaurants across the US. The day-fixed effects thus account for, e.g., seasonal effects or national holidays that affect the number of visits across all restaurant types.  $Chinese_i$  is a dummy variable that takes



**Figure 1: Relative Effect on Other Ethnic Cuisine Visits of a 0.01-Negativity Increase in US Newspaper Reporting on China**

the value of one if restaurant  $i$  offers Chinese food.  $\beta$  is the coefficient of interest and indicates how the political conflict, measured by *Negativity<sub>*t*</sub>*, in media reporting about China on day  $t$ , affects the number of visits to Chinese restaurants on day  $t$  relative to the average non-Chinese restaurant in the US.

#### Effect on Chinese Restaurants

Overall, our findings suggest that increases in political conflict with China have a statistically and economically meaningful impact on visits to Chinese restaurants in the US. Across all our specifications, the estimated effect  $\beta$  is statistically significant and negative, indicating that more political conflict with China is associated with a decline in visits to Chinese restaurants relative to the average non-Chinese restaurant

in the US. Assuming an increase in the week's average reporting negativity of 0.01 (corresponding to one additional negative word in each 100-word long paragraph), visits to Chinese restaurants decrease by 3.6% in our preferred specification. Our results are robust to including the control variables and estimating an array of alternative time series models. Further, the results are externally valid: applying our approach to the US-Mexico conflict during the Trump presidency yields qualitatively similar reductions in visits to Mexican restaurants after more political conflict with Mexico.

#### Effect on Other Ethnic Restaurants

The existence of ethnocentric tendencies suggests that consumers might decrease their

visits to foreign restaurants in general, even when the political conflict is only with one specific foreign country. Figure 1 shows that our data support this theory. We illustrate the estimated effects of a 0.01 increase in our measure of political conflict with China on other ethnic restaurants' visits. All estimated effects are highly significant, except for the small positive effect on Italian restaurants. Most strikingly, Figure 1 shows that our model estimates an almost 2.5% increase in visits to restaurants serving American food. The estimated effects for Mexican and Italian restaurants are comparatively small, with less than a 0.5% change in the number of visits. At the same time, our estimates for all other East-Asian restaurants are large and negative. This result suggests that US consumers might not always distinguish between Asian ethnic cuisines. Finally, Greek and Indian Food restaurants also exhibit sizable negative coefficients, although we consider these restaurant types as easily distinguishable from Chinese restaurants.

#### Conclusion

This study illustrates how smartphone location data can enable firms and researchers to answer empirical questions that have previously been hard to evaluate. These data provide information on daily visits to most locations – including commercial shops and restaurants – thus enabling researchers to estimate consumption reactions at scale and locations where the researcher would not have had data access under ordinary circumstances.

While real-world anecdotes and survey-based research suggest that political conflicts can have large effects on the consumption of foreign products, it is often difficult to quantify these effects separately from accompanying confounders such as, e.g., changes in prices and availability of products. We overcome this challenge based on novel data sources and demonstrate that an increased political conflict between the US and China substantially reduces visits to Chinese and other foreign restaurants. This finding suggests that consumers behave ethnocentric, which holds implications for policy makers and firms. Policy makers should communicate with that finding in mind during political conflicts. Additionally, firms should consider ethnocentrism as a relevant risk factor for product branding: products associated with a political conflict could suffer, even if they are not the subject of the conflict and are merely perceived as foreign.

#### References

**Sharma, S.; Shimp, T.; Shin, J.:**

Consumer Ethnocentrism: A Test of Antecedents and Moderators.

In: Journal of the Academy of Marketing Science, 23 (1994) 1, pp. 26–37.

**Shimp, T.; Sharma, S.:**

Consumer Ethnocentrism: Construction and Validation of the CETSCALE.

In: Journal of Marketing Research, 24 (1987) 3, pp. 280–289.

## Research Report

# Regulatory Impact Analysis in Case of Unstructured Data

REGULATORY IMPACT ANALYSIS (RIA) SERVES TO EVALUATE WHETHER REGULATORY ACTIONS FULFILL THE DESIRED GOALS. ALTHOUGH THERE ARE DIFFERENT FRAMEWORKS FOR CONDUCTING RIA, THEY ARE ONLY APPLICABLE TO REGULATIONS WHOSE IMPACT CAN BE MEASURED WITH STRUCTURED DATA. YET, A SIGNIFICANT AND INCREASING NUMBER OF REGULATIONS REQUIRE FIRMS TO COMPLY BY COMMUNICATING TEXTUAL DATA TO CONSUMERS AND SUPERVISORS. THEREFORE, WE DEVELOP A METHODOLOGICAL FRAMEWORK FOR RIA IN CASE OF UNSTRUCTURED DATA BASED ON TEXTUAL ANALYSIS AND APPLY IT TO A RECENT FINANCIAL MARKET REGULATION: MIFID II.

Benjamin Clapham

Jens Lausen

Micha Bender

Peter Gomber

### Introduction

Regulation is a fundamental tool for governments and policy makers to ensure customer and investor protection as well as market efficiency and integrity. In order to be effective and to achieve high regulatory quality, it is critical to ensure that regulatory adjustments and new regulations meet their desired objectives and result in the intended changes. The growing pace of technological progress and the increasing interdependencies between different financial regulations pose substantial challenges to policy makers and regulatory quality since the exact effects of a regulation are hard to assess.

For this purpose, policy makers and regulators around the globe conduct regulatory impact analysis (RIA) to evaluate whether regulatory actions meet the desired goals. Although there exist different guidelines and frameworks for conducting RIA (OECD, 1997; Radaelli, 2004), they are only applicable to regulations whose impact can be measured with structured and quantifiable data. Yet, an increasing and significant number of regulatory actions aim at or result in vast amounts of documents representing textual data that is hard to evaluate manually. To enable regulators and researchers to assess the impact of regulatory actions aimed

at unstructured data and to improve evidence-based policy making with the help of regulatory intelligence and RegTech-solutions, our paper (Clapham et al., 2023) develops a methodological framework for RIA in case of unstructured data. It builds on methods from textual analysis (TA) and natural language processing (NLP). We evaluate the RIA-framework and make use of it to assess the impact of a recent regulation: the changes in best execution requirements of the Markets in Financial Instruments Directive II (MiFID II) in Europe that are applicable since January 2018. These rule changes require investment firms, i.e., banks and brokers, to provide more informative best execution policies, which also should be easier to understand. In best execution policies, investment firms have to describe their processes of order handling and order routing to achieve the best possible result for their clients. Thus, these policies should enhance transparency for investors and protect them from potential downsides of the existing stock market fragmentation in Europe.

### Development of the Framework

Following the design science research paradigm (Hevner et al., 2004), we develop a framework for the analysis and evaluation of regulatory actions that result in unstructured data such as text documents. To create the framework, we build on existing RIA guidelines and on methods from TA and NLP. The RIA-framework provides detailed guidance as well as the required steps and tools to analyze the impact

of a regulation targeting at or leading to unstructured data in a systematic and largely automated manner. The framework (shown in Figure 1) can be applied as follows:

**Step 1:** Clearly describe the problem that the regulatory action wants to solve and identify the intended goals of the regulation.

**Step 2:** Identify the specific dimensions that are affected by the regulatory action. Dimensions in this context refer to the means (e.g., informativeness) that are targeted by the regulation in order to achieve the identified regulatory goal.

**Step 3:** Acquire the necessary data. The analysis and data acquisition approach for RIA in case of unstructured data differs dependent on whether a regulatory change or a new regulation is to be analyzed. In case of the revision of an already existing regulation, the affected objects, i.e., text documents, before as well as after the introduction of the regulatory change have to be collected. In case of the introduction of a new regulation, the collection of data before the regulatory change is mostly impossible since it often requires firms to publish new textual documents. Therefore, we propose a benchmark approach to assess the impact of new regulations. To collect appropriate benchmarks (e.g., textual data generated in similar regulatory areas), it is important to ensure comparability between affected objects and chosen benchmarks.

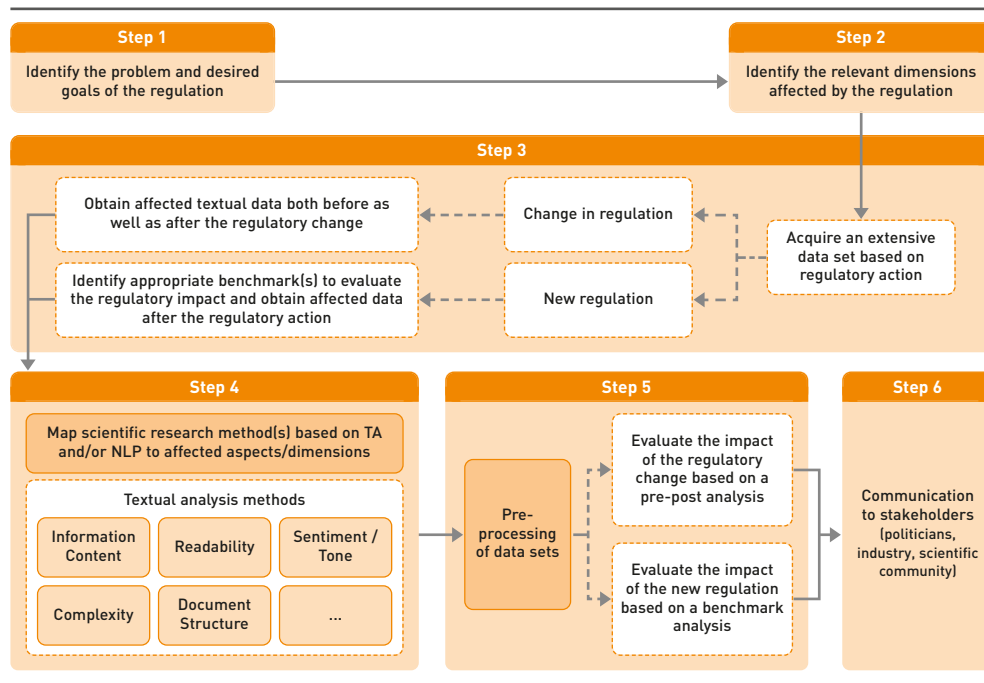


Figure 1: RIA-Framework for Unstructured Data

**Step 4:** Map appropriate scientific research method(s) to affected regulatory dimensions. The mapping of TA and NLP methodologies (see, e.g., Loughran and McDonald, 2016) to affected regulatory dimensions is a major component for an effective RIA in case of unstructured data and needs to be adapted for each specific use case.

**Step 5:** Preprocess data and evaluate the impact of the regulatory change or of a new regulation. This step comprises the actual analysis and evaluation of the regulation's impact. Depending on

the structure of the respective data (e.g., simple text files, documents in PDF format, textual information from web pages, XML, JSON), several preprocessing steps need to be performed to make the data machine-readable. Relevant textual information needs to be extracted and the handling of textual information from figures, tables, and lists needs to be determined. Depending on the methodology used for the analysis, further text cleaning steps such as removing stopwords or stemming need to be performed. For the assessment of the regulatory actions, the TA and NLP methods determined in

Step 4 are applied. Once these analyses are conducted, the results need to be compared, e.g., by using data visualization and statistical tests to evaluate whether changes in the analyzed measures can be observed and whether these changes correspond to the regulatory objectives.

**Step 6:** Communicate results. The final step of the framework represents the communication of the results of the RIA to relevant stakeholders, e.g., policy makers, regulators, consumer protection associations, and the scientific community, by publishing a policy white paper or a research report.

#### Empirical Evaluation: Best Execution Policies in MiFID II

To demonstrate its applicability, we use the RIA-framework to perform a regulatory impact analysis of the best execution requirements outlined in MiFID II. With more than 300 trading venues (as of September 2022), the European securities market is highly fragmented. Consequently, there is a large choice of venues for the execution of an order in a financial instrument. The selection of the appropriate trading venue by the investment firm for the customer order shall ensure the best possible result for the client ("best execution"). In this process, investment firms have to take into account a range of factors such as price, costs, speed, likelihood of execution and have to document their choices in best execution policies that are accessible to their clients. With MiFID II, which has to be applied since January 2018,

European authorities intend to improve best execution policies by requiring firms to specify and publish policies that are more *informative* and *comprehensible* in order to provide value to clients and to foster competition between investment firms.

We collect best execution policies of 50 German banks and brokers before as well as after the application of MiFID II. For a sample of 187 European best execution policies, we also perform a benchmark analysis and compare it to textual data generated in similar areas. In order to assess the *informational content* of the policies, we rely on three different measures: textual similarity, the percentage of boilerplate information, and the share of specific information in the documents. For the analysis of the policies' *comprehensibility*, we rely on two sets of measures: (1) readability measures such as the average number of words per sentence or the Fog index, which estimates the years of formal education a person needs to understand a text on the first reading; (2) textual complexity measures such as document length or the number of conditional statements in a document.

In the pre-post-analysis of German best execution policies, we find that all analyzed measures did not change or even worsened after MiFID II went live. Our results show that best execution policies became harder to read and more complex after MiFID II. Also, we find that the policies became more similar, which indicates that the information provided in these documents is less

useful for investors to differentiate between the providers of order execution services. Besides, the policies remain relatively unspecific and include a large share of boilerplate information. Most policies recite large parts of the regulation, which is neither informative for clients nor does it foster competition between brokers based on how they handle client orders to achieve best execution.

Looking at the European sample of best execution policies, we find additional evidence that MiFID II failed to achieve the desired goals. Although investment firms are obliged to differentiate between retail and professional clients

in their policies, the distributions across the different readability and complexity measures almost completely overlap (see Figure 2). Policies aimed at retail clients are as hard to read as policies aimed at professional clients. Based on the benchmark analysis, we find that best execution policies are among the most difficult and complex texts among the benchmarks. Figure 2 exemplarily shows that best execution policies are as difficult to read as companies' annual reports in 10-K filings and almost as difficult to read as European regulatory documents themselves. Spoken language, Wikipedia articles, and textbook chapters are noticeably easier to read than best execution

policies. Similar results exist regarding textual complexity. Based on the number of conditional statements, best execution policies belong to the most complex documents. Moreover, best execution policies use a large variety of vocabulary as measured by the number of unique bigrams. They are comparable to the management discussion section of annual reports. In summary, the analysis shows that best execution policies are not easy to understand as intended by the regulator but are highly complex documents that are difficult to read. The comprehensibility of best execution policies is similar to regulatory documents and companies' annual reports, which is way above what can be expected from retail clients.

### Discussion of the Results

More and more regulations aim at or result in huge numbers of textual documents. To enable researchers and regulators to assess whether regulatory actions have met the desired goals, this study develops a framework for RIA in case of unstructured data and applies it to the MiFID II requirements on best execution policies. We show that the regulation has not achieved its goal to increase informativeness and comprehensibility for customers. With this study, we pave the way for a largely untapped field of research within the RegTech literature, i.e., RegTech and decision support for regulators and policy makers in addition to the current main fields: compliance by firms and supervision by competent authorities. Although the framework is based on a use case from finan-

cial regulation, the framework should be applicable also to regulatory initiatives of other economic sectors as no step of the framework is unique to the financial industry. Rather, the framework represents a general principle to solve a class of real-world problems, i.e., conducting RIA in case of unstructured data.

### References

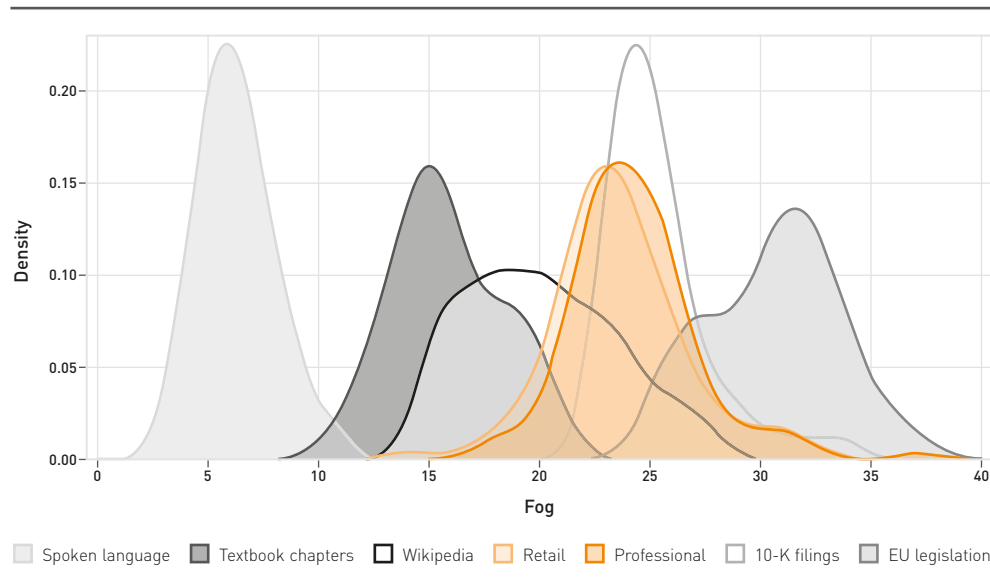
**Clapham, B.; Bender, M.; Lausen, J.; Gomber, P.:** Policy Making in the Financial Industry: A Framework for Regulatory Impact Analysis Using Textual Analysis. Forthcoming in: Journal of Business Economics.

**Hevner, A. R.; March, S. T.; Park, J.; Ram, S.:** Design Science in Information Systems Research. In: MIS Quarterly, 28 (2004) 1, pp. 75–105.

**Loughran, T.; McDonald, B.:** Textual Analysis in Accounting and Finance: A Survey. In: Journal of Accounting Research, 54 (2016) 4, pp. 1187–1230.

**OECD:** Regulatory Impact Analysis: Best Practices in OECD Countries, <http://www.oecd.org/regreform/regulatory-policy/35258828.pdf>, 1997.

**Radaelli, C. M.:** The Diffusion of Regulatory Impact Analysis – Best Practice or Lesson-Drawing? In: European Journal of Political Research, 43 (2004) 5, pp. 723–747.



**Figure 2: Comparison of the Readability of Best Execution Policies and Benchmark Documents Based on the Fog Index**



## Insideview

# Data Sharing: Don't Forget the Costs

INTERVIEW WITH STEFAN BENDER

**The discussion about data and its widespread use for the benefit of society continues to be widely discussed. Data sharing is on everyone's agenda. What needs to be considered?**

As the German government stated at its Digital Summit in 2022, data continues to be at the heart of digital change. Whether you like it or not, data influence our daily lives, such as production processes and supply chains or our consumer behavior. Therefore, the use of growing amounts of data is a key to innovation and sustainable growth. However, the cost/benefit calculation of data sharing should not be overlooked in the discussion.

**Cost/benefit analysis of data sharing? What does this mean?**

Data sharing only works if the benefits exceed the costs for all parties involved. The parties involved here are those from whom the data

originates, those who use the data and then us as a society. A different cost/benefit calculation has to be determined for each of these parties. For example, a data provider must comply with legal regulations – such as data protection – and technical challenges to fulfill data security and runs the risk of data misuse. These are clearly cost factors of data sharing. The benefits – e.g., through better data quality because of better documentation, increased use of the data or more relevant results from these data – must exceed the costs.

**Can you give a concrete example of this?**

Of course: through its Research Data and Service Center, the Bundesbank offers access to sensitive microdata for independent research. For the access, some conditions have to be fulfilled by the researchers, which may vary through the different data sets. If you want to communicate this information to the researchers in a systematic and machine-read-



Prof. Stefan Bender  
Head of Data Service Center  
Deutsche Bundesbank

able way, you quickly realize: the existing metadata schema describes the content of the data sets, but does not make it possible to systematically document data access rules. Therefore, we – together with colleagues from other central banks in our data sharing network "INEXDA" – have developed the so-called Annodata schema. It not only facilitates data access, but also increases the reproducibility of research results.

**Does such a schema also have further practical uses?**

Indeed, the machine-readability of Annodata schema was central. Therefore, we were able to develop our automated contract generator on the basis of this schema. That is why we could switch from the manual preparation of our contracts for data access to a software solution. That is digitalization in practice. However, the Annodata schema is only one measure among many by which we reduce the

costs of data access. For example, we also have tools for output control, because only absolutely anonymous results may be published.

**All in all, are we on the right track when it comes to data sharing?**

The EU Commission adopted the European Data Strategy in February 2020. This is intended to establish an internal European market for data. The EU wants to develop into one of the most secure, agile and attractive data areas in international comparison. On the one hand, data literacy should be broadly increased, but – while maintaining data protection and data security – the provision and access to data should also be made easier. This is the only way to shift the cost/benefit ratio of data sharing to greater benefits for all actors and to use the high potential that lies in the data for the benefit of society.

**Thank you for this interesting conversation.**

# Infopool

## News

### 20<sup>th</sup> Anniversary efl Conference Successfully Took Place on October 11<sup>th</sup>, 2022

The efl annual conference on “Democratization of Data Science and AI” covered a broad range of topics including machine learning, data cleaning, or state-of-the-art visualization techniques. Moreover, ethical aspects of AI were discussed and a perspective from Deutsche Bundesbank on the topic of data democratization was provided. The laudatory speech for 20 years efl was given by Prof. Dr. Joachim Wuermeling. The conference was organized by Prof. Dr. Carsten Binnig (TU Darmstadt) and his team: Nadja Geisler, Benjamin Hättasch, Benjamin Hilprecht, Adrian Lutsch, and Meghdad Mirabi. For more information, please visit: <https://www.eflab.de/annual-conference-2022>.

### Dr. Jens Lausen Received Prize for the Best Dissertation 2021/2022

Dr. Jens Lausen, former doctoral student at the Chair of e-Finance (Prof. Dr. Peter Gomber), was awarded the Dr. Mihael Foit and Elfriede Stich-Foit Foundation Prize (awarded for the first time) for the best dissertation at the Faculty of Economics and Business at Goethe University Frankfurt in the academic year 2021/2022. Congratulations!

### Dr. Cristina Mihale-Wilson Received Prize in the Management Disciplines

This year, the Schmalenbach Prize, endowed with EUR 10,000, was awarded to Dr. Cristina Mihale-Wilson for her dissertation. She is a post-doc at the Chair of Information Systems and Information Management of Prof. Dr. Oliver Hinz. In her dissertation “Contextual Antecedents and Consequences of Technology Adoption and Use”, she analyses contextual requirements and consequences of the adoption and use of new technologies such as AI-based assistance systems. Congratulations!

### Dr. Lennart Kraft Received “Dr.-Dietmar-Harting-Preis”

Dr. Lennart Kraft received the “Dr.-Dietmar-Harting-Preis” of the Erich-Gutenberg-Arbeitsgemeinschaft for his dissertation “Price of Privacy”. Among other things, he outlines that differential privacy is an alternative approach for preserving consumer privacy. Loosely speaking, differential privacy transforms some of the values that belong to an individual so that the individual can no longer be identified, but firms can still conclude about, e.g., the success of online advertising campaigns. Congratulations!

### Prof. Dr. Carsten Binnig was Program Chair for ACM SoCC 2022

SoCC’22 is the 13<sup>th</sup> edition of the annual ACM (association for computing machinery) symposium on cloud computing, the premier conference on cloud computing. It is co-sponsored by the ACM Special Interest Group on Management of Data (SIGMOD) and on Operating Systems (SIGOPS).

### The DFG Funds the Project “Understanding the Market Reaction to Audio Cues in Earnings Conference Calls” Supervised by Prof. Dr. Oliver Hinz

The research project aims to understand and design a profit-maximizing recommender system (PMRS), as well as examine its impact on online consumer behavior, the firm’s profit and network structure. At Goethe University, Maximilian Lowin and Moritz von Zahn are involved in the project.

## Selected efl Publications

### Carl, K. V.:

The Status-Quo of Companies’ Data Privacy and Security Communication: An Ethical Evaluation and Future Paths.

In: Proceedings of the 2022 INFORMATIK Conference; Hamburg, Germany, 2022.

### Clapham, B.; Bender, M.; Lausen, J.; Gomber, P.:

Policy Making in the Financial Industry: A Framework for Regulatory Impact Analysis Using Textual Analysis.

Forthcoming in: Journal of Business Economics.

### Clapham, B.; Haferkorn, M.; Zimmermann, K.:

The Impact of High-Frequency Trading on Modern Securities Markets – An Analysis Based on a Technical Interruption.

Forthcoming in: Business & Information Systems Engineering.

### El-Hindi, M.; Zhao, Z.; Binnig, C.:

Towards Decentralized Parameter Servers for Secure Federated Learning.

In: Proceedings of the 11<sup>th</sup> International Conference on Data Science, Technology and Applications (DATA); Lisbon, Portugal, 2022.

### Gassen, M.; Hättasch, B.; Hilprecht, B.; Geisler, N.; Fraser, A.; Binnig, C.:

Demonstrating CAT: Synthesizing Data-Aware Conversational Agents for Transactional Databases.

In: VLDB Endowment, 15 (2022) 12, pp. 3586–3589.

### Geisler, N.; Hättasch, B.; Binnig, C.:

Demonstrating Quest: A Query-Driven Framework to Explain Classification Models on Tabular. In: VLDB Endowment, 15 (2022) 12, pp. 3722–3725.

### Hättasch, B.; Bodensohn, J.-M.; Binnig, C.:

Demonstrating ASET: Ad-hoc Structured Exploration of Text Collections.

In: Proceedings of the 2022 International Conference on Management of Data (SIGMOD); Philadelphia (PA), US, 2022.

### Hilprecht, B.; Binnig, C.:

Zero-Shot Cost Models for Out-Of-The-Box Learned Cost Prediction.

In: VLDB Endowment, 15 (2022) 11, pp. 2361–2374.

### Matthe, M.; Ringel, D.; Skiera, B.:

Mapping Market Structure Evolution.

Forthcoming in: Marketing Science.

### Skiera, B.; Yan, S.; Daxenberger, J.; Dombois, M.; Gurevych, I.:

Information Seeking Argument Mining: A Step Towards Identifying Reasons in Textual Analysis to Improve Services.

Forthcoming in: Journal of Service Research.

### Zacharias, J.; von Zahn, M.; Chen, J.; Hinz, O.:

Designing a Feature Selection Method based on Explainable Artificial Intelligence.

Forthcoming in: Electronic Markets.

For a comprehensive list of all efl news postings and efl publications see: <https://www.eflab.de>

# Infopool

## RESEARCH PAPER:

### AUGMENTING DECISION MAKING VIA INTERACTIVE WHAT-IF ANALYSIS

The fundamental goal of business data analysis is to improve business decisions using data. To discover the relationship between data attributes hypothesized to be drivers and those corresponding to key performance indicators (KPIs) of interest, business users currently need to perform lengthy exploratory analyses. However, the increasing complexity of datasets combined with the cognitive limitations of humans makes it challenging to carry over multiple hypotheses, even for simple datasets. Hence, this paper argues for four new functionalities that are necessary to enable business users to interactively learn and reason about the relationships (functions) between sets of data attributes thereby facilitating data-driven decision-making. Moreover, the paper evaluates the system through three business use cases – marketing mix modeling, customer retention analysis, and deal closing analysis – and reports on feedback from multiple business users.

**Gathani, S.; Hulsebos, M.; Gale, J.; Haas, P. J.; Demiralp, Ç.**

**In: Proceedings of the 12<sup>th</sup> Annual Conference on Innovative Data Systems Research (CIDR); Santa Cruz (CA), US, 2022.**

## RESEARCH PAPER:

### WHAT MOVES STOCK PRICES? THE ROLES OF NEWS, NOISE, AND INFORMATION

The issue of what drives stock price movements is a fundamental question in finance with implications for understanding risk, informational efficiency, and asset pricing. To better understand the stock-return generating process, the paper develops a return variance decomposition model to distinguish the roles of different types of information and noise in stock price movements. Thereby, it disentangles four components: noise, private firm-specific information revealed through trading, firm-specific information revealed through public sources and market-wide information. Overall, the authors find that 31% of the return variance is from noise, 24% from private firm-specific information, 37% from public firm-specific information and 8% from market-wide information. The results are in line with the general development of a declining share of noise and an increasing share of firm-specific information that shapes prices since the mid-1990s, being consistent with increasing market efficiency.

**Brogaard, J.; Nguyen, T. H.; Putnins, T. J.; Wu, E.**

**In: The Review of Financial Studies, 35 (2022) 9, pp. 4341–4386.**

## efl insights

The efl publishes the insights in the form of a periodic newsletter which appears two times a year. Besides a number of printed copies, the efl insights is distributed digitally via E-mail for reasons of saving natural resources. The main purpose of the efl insights is to provide latest efl research results to our audience. Therefore, the main part is the description of two research results on a managerial level – complemented by an editorial, an interview, and some short news.

For receiving our efl insights regularly via E-mail, please subscribe on our homepage [www.eflab.de](http://www.eflab.de) (> NEWS > SIGN UP EFL INSIGHTS) as we need your E-mail address for sending the efl insights to you. Alternatively, you can mail your business card with the note “efl insights” to the subsequent postal address or send us an E-mail.

**Prof. Dr. Peter Gomber**

**Vice Chairman of the efl – the Data Science Institute**

**Goethe University Frankfurt**

**Theodor-W.-Adorno-Platz 4**

**D-60629 Frankfurt am Main**

**[insights@eflab.de](mailto:insights@eflab.de)**

**Further information about the efl is available at  
[www.eflab.de](http://www.eflab.de).**



The efl – the Data Science Institute is a proud member of the House of Finance of Goethe University, Frankfurt.  
For more information about the House of Finance, please visit [www.hof.uni-frankfurt.de](http://www.hof.uni-frankfurt.de).

THE EFL – THE DATA SCIENCE INSTITUTE IS AN INDUSTRY-ACADEMIC RESEARCH PARTNERSHIP BETWEEN FRANKFURT AND DARMSTADT UNIVERSITIES AND PARTNERS DEUTSCHE BÖRSE GROUP, DZ BANK GROUP, AMERICAN EXPRESS, FACTSET, AND FINANZ INFORMATIK LOCATED AT THE HOUSE OF FINANCE, GOETHE UNIVERSITY, FRANKFURT.

For further  
information  
please  
contact:

Prof. Dr. Peter Gomber  
Vice Chairman of the  
efl – the Data Science Institute  
Goethe University Frankfurt  
Theodor-W.-Adorno-Platz 4  
D-60629 Frankfurt am Main

**Phone** +49 (0)69 / 798 - 346 82  
**E-mail** [gomber@wiwi.uni-frankfurt.de](mailto:gomber@wiwi.uni-frankfurt.de)

**Press contact**  
**Phone** +49 (0)69 / 798 - 346 82  
**E-mail** [presse@eflab.de](mailto:presse@eflab.de)

**or visit our website**  
<http://www.eflab.de>