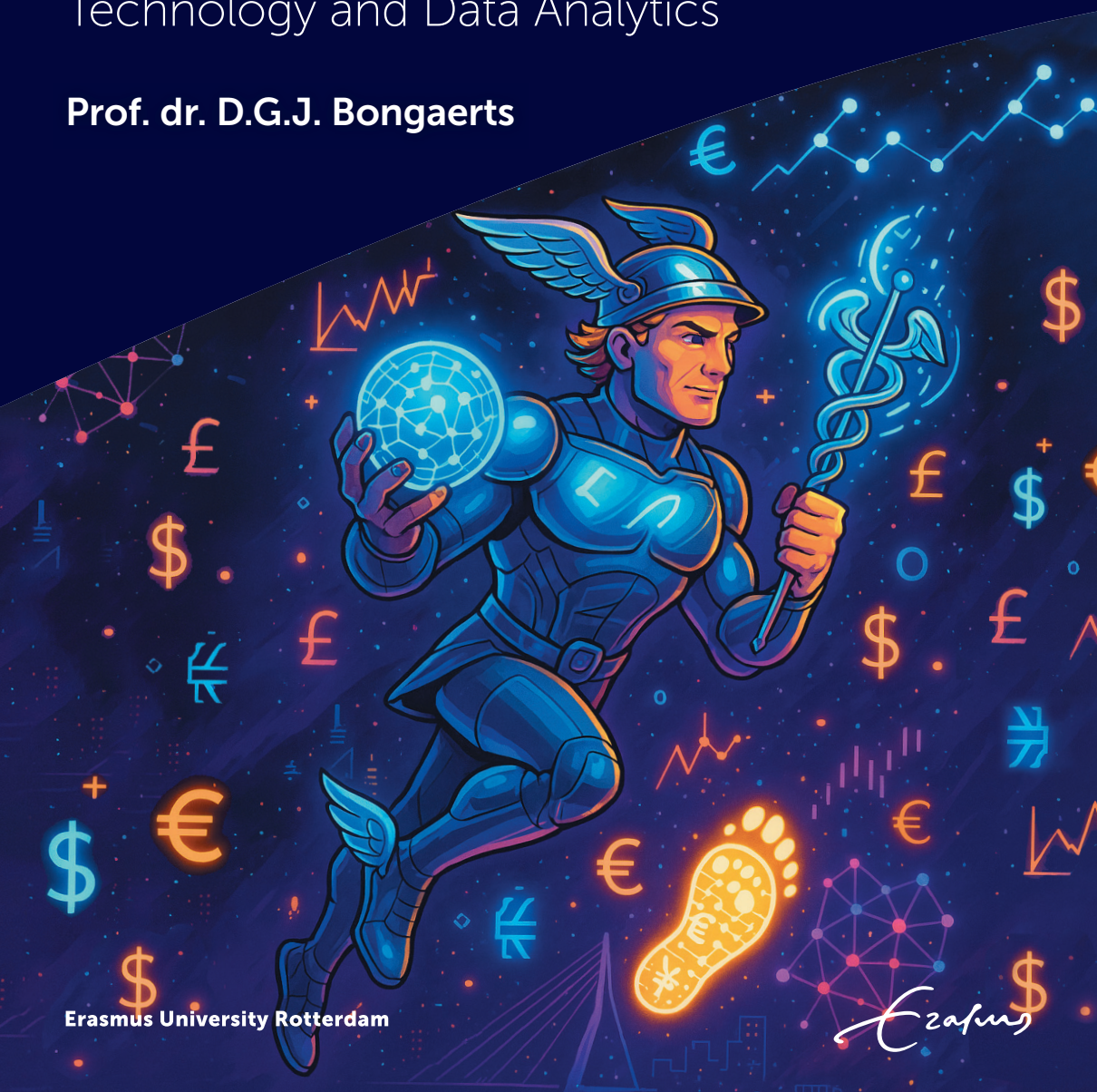


The Financial Sector as a Data Trove

Inaugural Address for the Chair in Financial
Technology and Data Analytics

Prof. dr. D.G.J. Bongaerts



The Financial Sector as a Data Trove

Inaugural Address for the Chair in Financial
Technology and Data Analytics

Erasmus Research Institute of Management - ERIM

The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus Universiteit Rotterdam
Internet: www.erim.eur.nl

<https://pure.eur.nl>

Inaugural Addresses Research in Management Series

Reference number EIA-2025-90-F&A

ISBN 978-90-5892-737-8

© 2025, Prof. dr. D.G.J. Bongaerts

Design

PanArt creatie en communicatie, www.panart.nl

Print

OBT bv, www.obt.eu

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

This publication (cover and interior) is printed on certified FSC® paper PlanoSuperior.



The Financial Sector as a Data Trove

Inaugural Address for the Chair in Financial
Technology and Data Analytics

Prof. dr. D.G.J. Bongaerts

Address delivered at the occasion of accepting the appointment as
endowed professor of Financial Technology and Data Analytics on behalf of
Erasmus University Rotterdam, at the Rotterdam School of Management
Erasmus University Rotterdam on 13th, June, 2025.

Rotterdam School of Management
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
E-mail: dbongaerts@rsm.nl
<https://www.rsm.nl/people/dion-bongaerts/>

Inaugural Lecture on 13th June 2025. Prof. dr. D.G.J. Bongaerts

The Financial Sector as a Data Trove

Chair in Financial Technology and Data Analytics

Abstract

In my inaugural lecture as Chair of Financial Technology and Data Analytics at the Rotterdam School of Management, Erasmus University, I start from the insight that today's financial systems are purely based on administrative constructs. Data, and the way data is handled, are therefore of crucial importance within these systems. Such data is therefore particularly well-suited for analyzing economic behavior. I provide various examples of this. In addition, I demonstrate how heterogeneity among market participants in their ability to conduct such analyses can, in turn, influence economic outcomes such as liquidity and market quality in equity markets. I also relate the storage and analysis of financial data to modern innovations such as blockchain, synthetic data, and AI models, and illustrate how these technologies can help researchers unlock data and measure abstract and qualitative concepts. I conclude with surprising operational applications of these technologies within the university.

The Financial Sector as a Data Trove

Leerstoel Financiële Technologie en Data Analyse

Samenvatting

In mijn inaugurele rede bij de aanvaarding van de leerstoel Financiële Technologie en Data Analyse aan de Rotterdam School of Management, Erasmus Universiteit, begin ik bij het inzicht dat de hedendaagse financiële systemen puur gebaseerd zijn op administratieve constructen. Data en de manier waarop met data wordt omgegaan zijn dan ook van cruciaal belang binnen deze systemen. Deze data zijn bij uitstek geschikt om economisch gedrag te analyseren. Ik geef hier diverse voorbeelden van. Ook laat ik zien hoe verschillen tussen marktpartijen in het vermogen dit soort analyses te doen zelf weer van invloed kunnen zijn op economische uitkomsten zoals bijvoorbeeld liquiditeit en markt kwaliteit in aandelenmarkten. Ik relateer de opslag en analyse van financiële data ook aan moderne innovaties zoals blockchain, synthetische data, en AI-modellen en illustreer hoe deze technieken onderzoekers kunnen helpen in het ontsluiten van data en het meten van abstracte en kwalitatieve concepten. Ik sluit af met verrassende operationele toepassingen van deze technologieën binnen de universiteit.

Table of contents

Inaugural address Dion Bongaerts11

Oratie Dion Bongaerts23

References/Referenties.....34

Inaugural address Dion Bongaerts

Chair in Financial Technology and Data Analytics

The financial sector as a data trove

Welcome today at my inaugural address for my chair in Financial Technology and Data Analytics. I am honored and grateful to be appointed as a full professor on this chair and for all of you coming over to listen to what I have to say on these topics. Today, I will be providing an, I hope, rather intuitive introduction to my field that is on the intersection of finance, econometrics, data science, and software engineering (in that order of importance). If this sounds intimidating, I illustrate my talk visually in two ways. For the more consultancy-oriented people in the audience, I have flow charts on the right-hand side of the slides. For whoever finds that too boring, I've tried to turn this talk into a science-fiction comic on the left-hand side.

A couple of years ago I was designing a lecture on modern payment systems that I use in various programs when an important insight struck me:

The whole financial sector these days is essentially one big administrative system!

To see, just think about your own account balance. Banks keep a record of our savings and borrowings and in turn, central banks keep a balance of their savings, borrowings, and collateral. Similarly, brokers keep track of our investments (which are nothing more than records in their administrative systems) and in turn, their associated holdings are administered by so-called CSDs. This means that money, securities, crypto, and debt are purely artificial administrative constructs. Yet, as we all know, these are crucially important in today's society and economy. This is a key take-away from that class and a related 2Tokens masterclass on token economics that I regularly teach.

Because of the aforementioned importance of financial administrations, we find footprints of the behavior and decisions of private individuals or corporates in the administrative systems of financial institutions like banks. As a result, the data in these systems are a treasure trove of information if one wants to study and understand the behavior of economic agents such as companies or private individuals as ourselves (Figure 1).

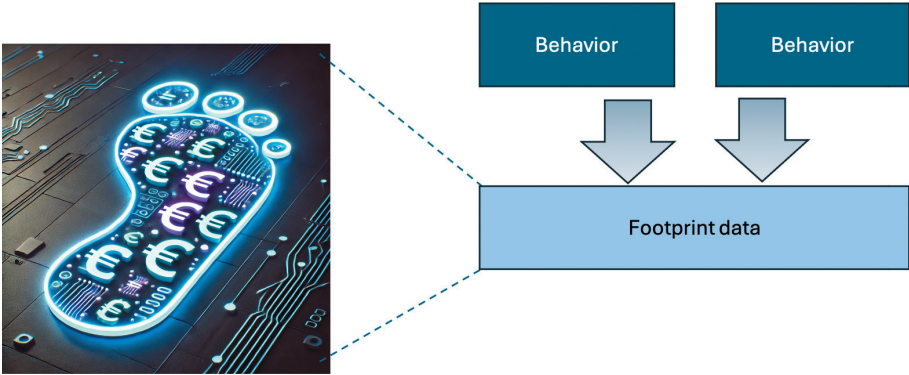


Figure 1.

It is important to realize that the information contained in these data are mere footprints and may fail to fully represent the economic and non-economic actions behind these footprints, just like your own footprint is not fully uninformative about the color of garment you wear. Therefore, if we want to extract information about behavior from these data, we need to use models and statistical techniques to basically try and reconstruct the underlying actions or motivations (e.g., from the direction of the footprint) or key features of the decision makers from the data (e.g., from the shape or depth of the footprint). This process is called (financial) data analysis (Figure 2).

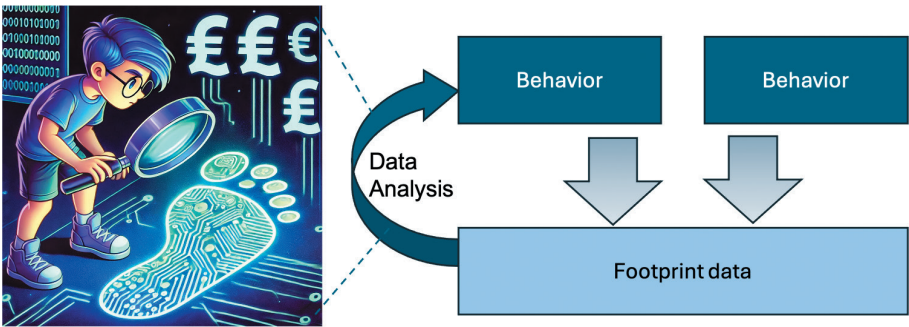


Figure 2.

A natural application of studying the footprints in financial data is a project together with Mathijs van Dijk, Darya Yuferova, Dominik Roesch, and Dick Roll that was published in *Management Science*. In the study, we explore the anatomy and propagation of financial crises, and in particular stock market crashes. For this, we use **32.5 bln** datapoints consisting of intra-day equity market quotes and trades from all over the globe. Specifically, we wanted to understand the nature of stock market crashes and how they spread across markets. The detailed and global coverage of our data allowed us to study stock market crashes at high frequency as they unfold. Still, because we for example cannot observe the identity of various traders, our identification is to a certain degree indirect and dependent on models. We did find that most stock market crashes are 1.) local and 2.) information driven. We find only very limited evidence for problems such as reduced liquidity due to a crash leading to amplification or contagion effects.

In the previous example, it was possible to employ an economic model to directly map the footprint of various crises to properties of these crashes. Yet sometimes one does not have this luxury and in order to study a specific phenomenon closely enough, such a model could be insufficient. In those cases, one could resort to advanced algorithms (typically in combination with a super-computer) to try and reverse engineer what actually happened at a more detailed level. A subsequent analysis can then generate insights from these “enhanced data” (Figure 3).

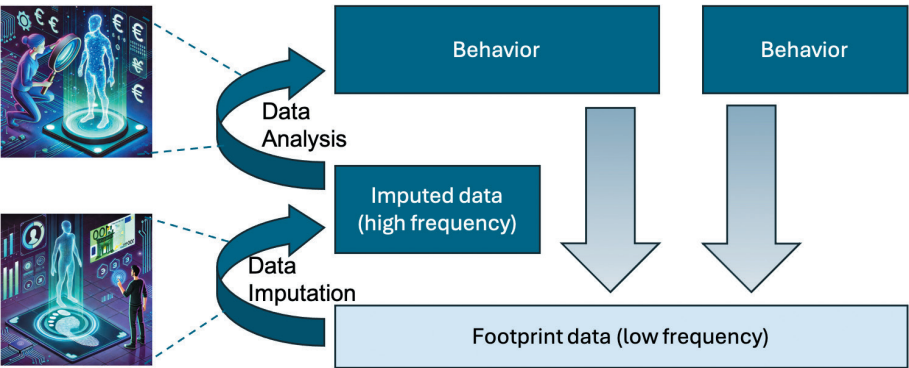


Figure 3.

A good example of using rather coarse data to reverse engineer what actually happened is a recent project together with Jean-Paul van Brakel and Mathijs van Dijk. In the project, we want to obtain the trading behavior of mutual funds in each position on each day. Unfortunately, funds only report their individual positions at best every month and sometimes only every quarter. Yet, funds are required to report some other data, namely aggregate portfolio returns every day (so not position by position). We use these limited data in combination with an advanced algorithm and an enormous amount of supercomputer capacity to impute the trades of every fund on every day (a normal computer would need to work non-stop for about 14 years to yield the same result). Of course, we still need to show that the method works, which we do using a proprietary dataset for which we observe daily positions and trades. Armed with our estimates, other researchers have shown that many mutual funds engage in window dressing with regards to sustainable investments (Parise and Rubin, 2025). Of course, our hope is that our research like this, in an indirect way leads to more transparency and accountability of mutual fund managers.

Given the earlier discussion on the importance of financial administrations, one can imagine that the way these administrations are maintained is of crucial importance. Ideally, such administrative systems should be easy to work with, have high capacity, and still be secure. The occasional disruption in payment systems shows the mayhem that can ensue when such a system fails.

Most professional-grade financial databases these days are organized as standard, centralized SQL databases. Typically, one party that enjoys a high level of trust, such as a bank, insurer, or government is in charge of the system (Figure 4). This is how our bank accounts operate. Yet, in 2009 a new technology arose to securely manage a database, namely a blockchain. It was brought to the public's attention with the most successful use case so far, namely the crypto-currency bitcoin.

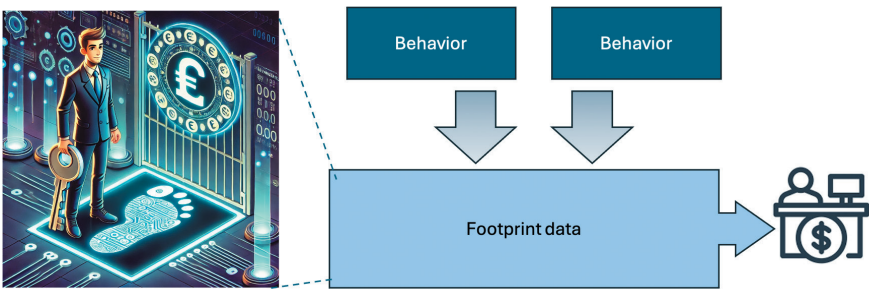


Figure 4.

While a centralized database requires trust in the administrator of that database, this is not true for a blockchain. In a blockchain, data is redundantly recorded by many different users that all check one another using cryptography (Figure 5). Hence, as long as the majority of users is honest or at least does not collaborate with one another to disadvantage the rest of the users, the data in this decentralized database is secure. Some blockchains can even run computer code, turning them into some kind of world supercomputer.

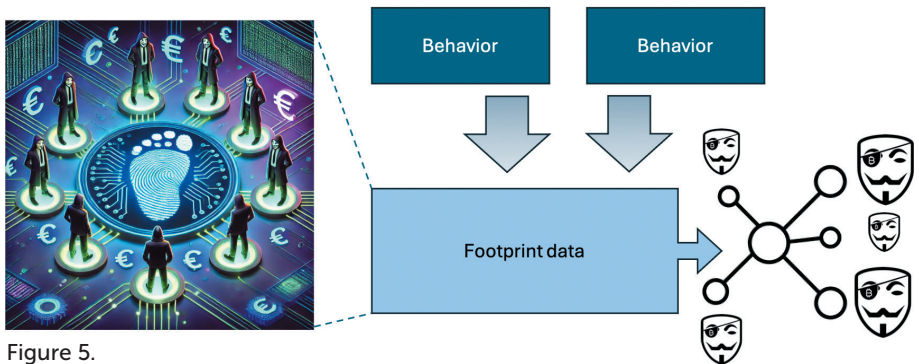


Figure 5.

Such a blockchain is an interesting new technology to store data and can be applied not only to crypto currencies, but also to other value objects like voting rights. One of the applications that have been built are so called DAOs, organizations or digital firms that only exist as an administration on the blockchain. The decision process of these DAOs is organized through user voting. Together with Dan Liebau, Thomas Lambert, and Peter Roosenboom, we study whether such a system of voting on economic decisions leads to more democratic decisions (i.e., involving everyone with the right to vote) than voting in shareholder meetings of traditional firms in which institutional investors such as pension funds typically dominate, and small shareholders do not bother to vote. We find that many users seem to indeed appreciate decentralization. Yet, there are large financial parties that keep up appearances of decentralization by letting others cast their votes. The only problem is that they have such tight control over the parties voting on their behalf that these few parties dominate the DAO completely.

Finally, having access to data, being able to process data more effectively, or having better infrastructure creates informational and technological advantages that are likely unequally distributed in society. The effect of such technology on economic outcomes can also be studied, both empirically as well as in the context of theoretical models.

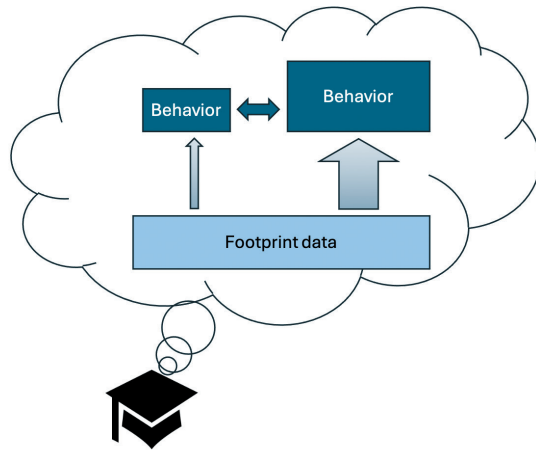


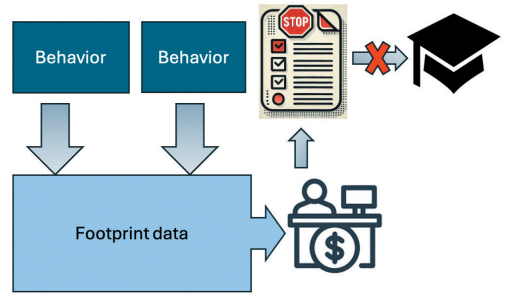
Figure 6.

One article in which the effect of advantages in information and/or technological infrastructure is studied from a theoretical perspective is a recent paper that I published together with Mark van Achter in the *Journal of Financial Economics*. In the study, we put forward a theoretical model of a securities market in which technologically advanced parties compete with less advanced parties in the context of equities trading (Figure 6). These markets operate at extremely high speeds (data are time-stamped to the nano-second) and so-called high-frequency traders have taken up a large part of this market. These high-frequency traders use advanced algorithms to predict whether the next person that wants to trade with them is better informed or not, which matters for their profitability. Hence, these high-frequency traders are typically better informed than other parties in the market, but their technology may be more expensive to operate. In the study, we investigate whether the availability of such technology is good for markets and society or not. We find that the advanced parties start to dominate due to their informational advantage. This is true even if their technology is more costly and less efficient to operate (and hence they should optimally not participate). Moreover, they tend to shun the market in a more correlated fashion as their technology generates warnings at the same time, leading to less stable markets. As a result, such technological innovations in information technology may leave markets, and thereby society, on aggregate worse off.

I often get the question of how I see this field evolving going forward. One of the constraints my field tends to hit over and over again is a lack of high-quality data for investigating specific research questions. The problem is typically not that the data is non-existent. The problem is that whoever owns the data does not want to or is not allowed to share those data for academic purposes (Figure 7). Privacy concerns play a sizable role in such considerations and rightfully so. For this reason, I am working on a new project, together with a TU-Delft originated startup called BlueGen on exploring the potential and security of synthetic data in finance.



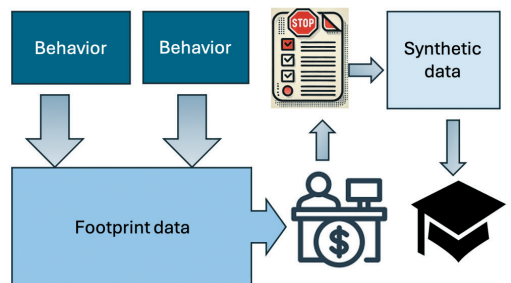
Figure 7.



Synthetic data are derived data with the same statistical properties as the data they are derived from. Yet, if done well, they are sufficiently different so that the original data cannot be reconstructed. Understanding the potential, limitations and trade-offs in synthetic data are a first step in inducing external parties like banks, insurers, and other institutions to share data. Yet, generating accurate but sufficiently secure synthetic data is not straightforward, and current algorithms are not tailored to specific properties that are found in many finance datasets. In this project we are exploring the difficulties synthetic data generators run into when faced with finance datasets and try to improve these generators. Armed with these results and improvements, our hopes are that students and researchers alike will be faced with much more willing compliance desks when asking corporate partners for proprietary data for academic research (Figure 8).



Figure 8.

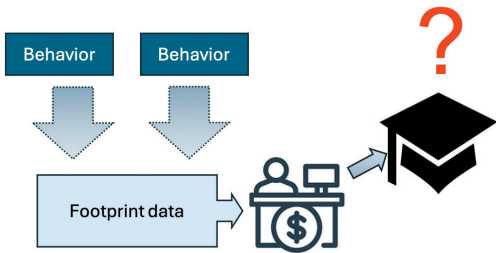


The software that generates synthetic data is a form of generative artificial intelligence or GenAI. Of course, this inaugural address would not be complete without me saying anything about large language models such as the ones employed by ChatGPT. I spoke before about our financial system being in essence a large administrative system of rights and obligations. I also explained that the footprints these data represent can be analyzed to better understand which actions were taken, why, by whom, and to what purpose. Yet, in many cases, other data is available as well in the form of rather unstructured text. Large language models are revolutionary in the sense that they learn the meaning of texts not because a human tells them what each sentence means, but rather because they can make associations with patterns in surrounding texts. Of course, one would need to feed the model tremendous amounts of textual data to get decent accuracy, but with the entire internet available, this is becoming less and less of an issue.

In a university, the introduction of Large Language Models (LLMs) and ChatGPT in particular was very noticeable. Our students were the so-called early adopters when they discovered the power of ChatGPT in doing their assignments. But with it, we ourselves as researchers also found an incredibly useful tool in our toolbox. This tool allows us to measure aspects of companies, securities, or individual that are not captured well with the financial footprint data (Figures 9 and 10). An example of such a project is the project together with my colleagues from accounting, Erik Peek and Stephan Kramer as well as one of our students, Pepijn de Kruijf. In the project, we explore the role of equity market analysts in disciplining management to increase firm value instead of consuming perks on their company's expense.



Figure 9.



Specifically, we are interested whether analysts with a more critical attitude discipline managers better. The problem is that a critical attitude is very hard to measure in quantitative data. Yet, these analysts make various formal and informal textual and spoken disclosures. Using a large language model, we then try to measure the degree of criticalness of each analyst. While still ongoing, we see this project as a promising way in utilizing Large Language Models in making the unmeasurable measurable and hope to apply the same technique elsewhere.

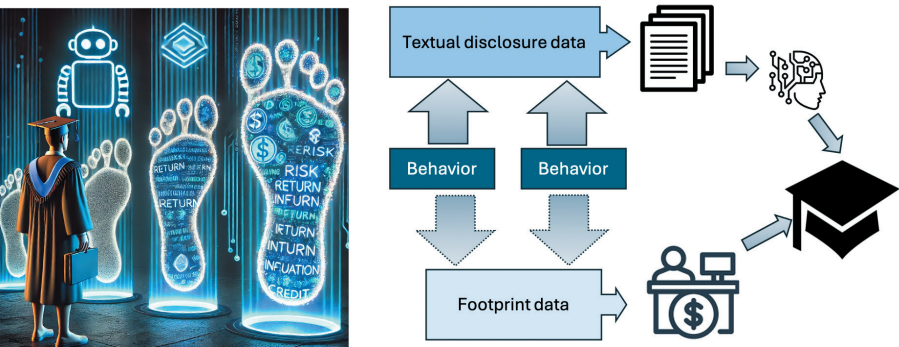


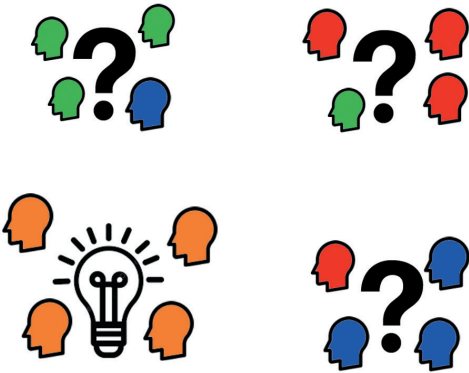
Figure 10.

The ability of algorithms to process unstructured text is something that is also useful in many contexts other than research. One such context is the mission of the university and business school in which diversity plays an important role. Diversity has been shown in various contexts to lead to better outcomes and decision-making processes, among others due to diversity of thinking. Yet, diversity of thinking is hard to measure, and hence, we often see the path of least resistance by focusing on the coarse, easily measured, and sometimes even politically hijacked, tick box dimensions of ethnicity and gender.

In my teaching, I did experience a diversity issue in terms of an imbalance of skills in student groups for group assignments. I would typically see talent cluster together and groups of less talented students complain about the difficulty of the assignment. Students would also not learn much from one another (Figure 11). Obviously, the standard tick-box approach to diversity would fail here, so I had to outsmart the problem.



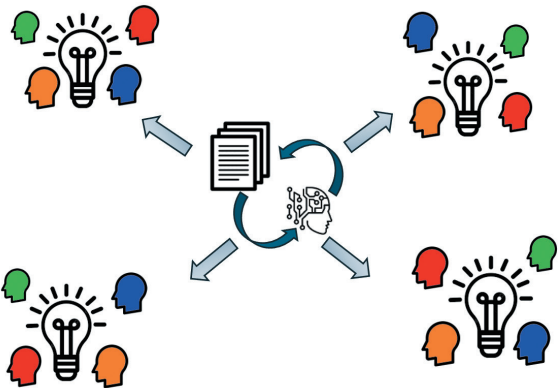
Figure 11.



For that reason, I asked students to write their own profile and let an algorithm cluster similar students together into equally sized clusters. I would then form student teams by taking one student from each cluster. Following this group formation method, there was a sharp drop in the number of complaints about assignment difficulty and students learned more from one another (Figure 12).



Figure 12.



It is the challenges in these research projects, but also the ability to innovate in teaching like I just described that make this job extremely interesting and satisfying and I am grateful to be able to do this job. As mentioned before, I am also extremely grateful for your presence and attention today and in a few moments, I will give you my token of appreciation in the form of a crypto coin that commemorates your attendance to this event. You can scan the QR code with your phone and add it to your digital wallet. Note, this is just an administration, the coin is administered on a blockchain, you won't magically find a physical coin in your pocket.

This token of appreciation represents among others the gratitude I feel towards the people that fostered my academic and personal growth. I have been lucky to work in engaging and stimulating environments at ABN-AMRO, the University of Amsterdam, Erasmus University, the various places in which I did research visits, and the various external partners that I have worked with and still work with. Some of my colleagues and contacts have gone to great lengths to support me and help me grow and some have become good friends in the process. It warms my heart to see so many of you present today.

In addition to a nurturing professional environment, I also praise myself happy with a group of close friends that I can share happy and sad moments with as well as views on the world, politics, family, love, sports, food, and wine.

Of course, I'd also like to thank the people who are most special to me for their enduring love and support. My dear sister and brother-in-law Erik-Jan, my wonderful parents - and in particular my mom who has a near telepathic ability of understanding me - and of course my loving wife and amazing daughter Niki.

Ik heb gezegd.

Oratie Dion Bongaerts

Leerstoel Financiële Technologie en Data Analyse

De Financiële sector: Een schatkist aan data

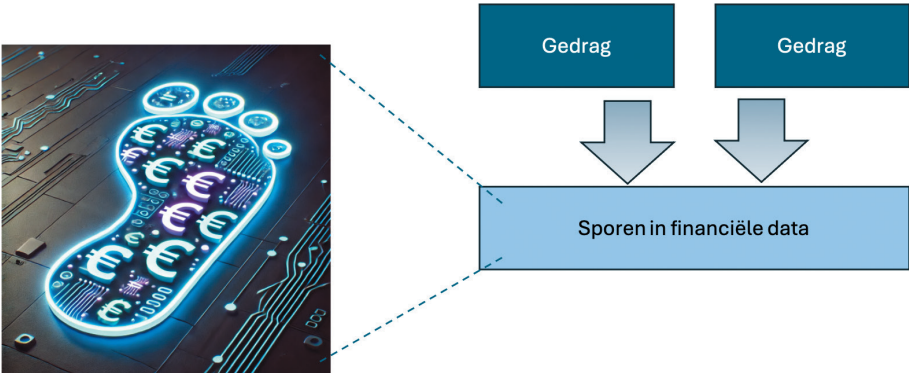
Welkom bij mijn oratie voor de leerstoel in Financiële Technologie and Data-Analyse. Ik voel me vereerd en ben dankbaar dat ik als hoogleraar op deze leerstoel ben benoemd en dat jullie zijn gekomen om te luisteren naar wat ik over dit onderwerp te vertellen heb. Vandaag geef ik een, hopelijk, vrij intuïtieve introductie tot mijn vakgebied, dat zich bevindt op het snijvlak van finance, econometrie, data science en software engineering (in die volgorde). Mocht dit intimiderend klinken: ik illustreer mijn verhaal visueel op twee manieren. Voor het meer consultancy-georiënteerde publiek heb ik stroomdiagrammen aan de rechterkant van de slides. Voor degenen die dat te saai vinden, heb ik geprobeerd van dit verhaal een sciencefiction-strip te maken aan de linkerkant.

Een paar jaar geleden was ik bezig met het ontwerpen van een college over moderne betaalsystemen dat ik in verschillende programma's gebruik, toen ik opeens tot een belangrijk inzicht kwam:

De hele financiële sector is tegenwoordig in essentie één groot administratiesysteem!

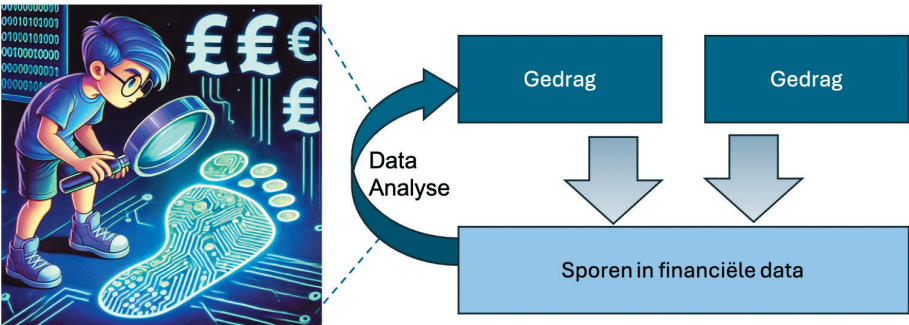
Denkt u bijvoorbeeld aan uw eigen bankrekening. Banken houden een administratie bij van onze spaartegoeden en leningen, en centrale banken houden op hun beurt weer een administratie bij van hun tegoeden, leningen en onderpanden. Ook brokers registreren onze beleggingen (die niet meer zijn dan administratieve entries in hun systemen) en hun bijbehorende posities worden weer bijgehouden door zogenaamde CSD's. Dit betekent dat geld, effecten, crypto en schuld puur kunstmatige, administratieve constructen zijn. Toch zijn ze, zoals we allemaal weten, van cruciaal belang in onze economie en maatschappij. Dit is een belangrijke les uit dat college en ook uit een gerelateerde 2Tokens-masterclass over token economie die ik regelmatig geef.

Vanwege het belang van financiële administraties vinden we sporen van het gedrag en de keuzes van individuen en bedrijven in de administratiesystemen van financiële instellingen zoals banken. Hierdoor vormen de gegevens in deze systemen een ware schatkist of goudmijn als men het gedrag van economische agenten wil bestuderen en begrijpen (Figuur 1).



Figuur 1.

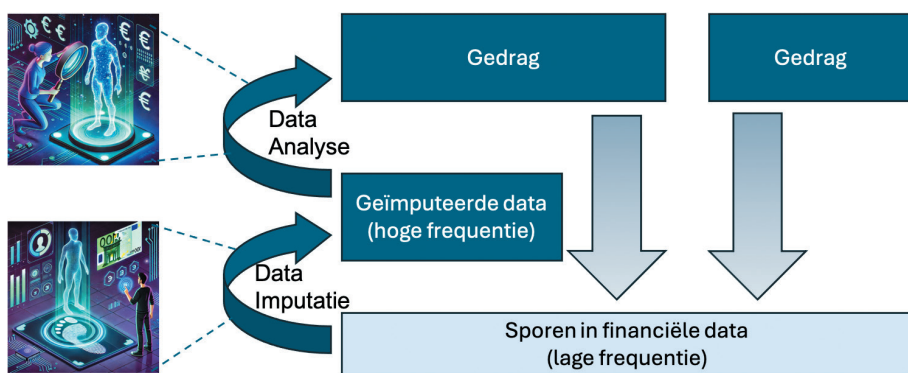
Het is belangrijk te beseffen dat de informatie in deze gegevens slechts sporen zijn en dat deze sporen mogelijk niet een volledig beeld geven van de economische of niet-economische acties die eraan ten grondslag liggen – net zoals je eigen voetafdruk weinig zegt over de kleur van je kleding. Als we dus informatie over gedrag willen afleiden uit deze data, moeten we modellen en statistische technieken gebruiken om de onderliggende handelingen of motivaties te reconstrueren (bijv. uit de richting van de voetafdruk), dan wel de kenmerken van de besluitvormers (bijv. uit de vorm of diepte van de afdruk). Dit proces heet (financiële) data-analyse (Figuur 2).



Figuur 2.

Een natuurlijke toepassing van het bestuderen van deze sporen in financiële data is een project dat ik samen uitgevoerd heb met Mathijs van Dijk, Darya Yuferova, Dominik Rösch en Dick Roll, dat is gepubliceerd in *Management Science*. In deze studie analyseren we de anatomie, verspreiding, en verloop van financiële crises, in het bijzonder beurscrashes. Hiervoor gebruikten we **32,5 miljard** datapunten bestaande uit binnen de dag geobserveerde beurskoersen en aandelentransacties van over de hele wereld. We wilden begrijpen hoe beurscrashes ontstaan en hoe deze zich verspreiden tussen markten. Dankzij de gedetailleerde en wereldwijde dekking van onze data konden we beurscrashes op hoge frequentie bestuderen op het moment dat ze zich ontvouwen. Toch is onze identificatie tot op zekere hoogte nog indirect, aangezien we bijvoorbeeld niet de identiteit van handelaren kunnen waarnemen en dus afhankelijk zijn van modellen. We vonden dat de meeste crashes lokaal en informatie-gedreven zijn. We vonden slechts beperkt bewijs voor problemen zoals opdrogende liquiditeit in het verergeren of verspreiden van crashes.

In het vorige voorbeeld konden we met behulp van een economisch model direct de sporen van crises koppelen aan eigenschappen van de crashes. Maar soms is zo'n model niet voldoende. In die gevallen kan men gebruikmaken van geavanceerde algoritmes (meestal met behulp van een supercomputer) om te proberen te achterhalen wat er op detailniveau precies is gebeurd. Met een daaropvolgende analyse kan men dan inzichten verkrijgen uit deze 'verrijkte data' (Figuur 3).

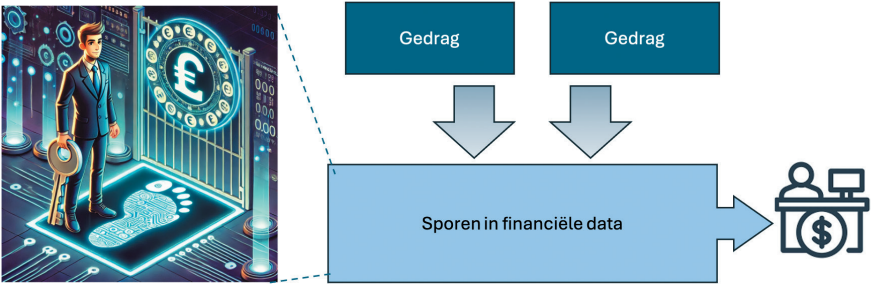


Figuur 3.

Een goed voorbeeld van het gebruik van beperkte data om gedrag te reconstrueren is een recent project samen met Jean-Paul van Brakel en Mathijs van Dijk. In dit project willen we het handelsgedrag van beleggingsfondsen op dag basis en per positie afleiden. Helaas rapporteren fondsen hun posities slechts maandelijks of zelfs maar eens per kwartaal. Wel zijn zij verplicht om dagelijks het rendement van de totale portefeuille te rapporteren. Wij combineren deze beperkte data met een geavanceerd algoritme en de enorme rekenkracht van een supercomputer (een gewone computer zou 14 jaar non-stop bezig zijn met deze berekeningen) om dagelijkse transacties per fonds te imputeren. We valideren onze methode met een dataset waarin we de dagelijkse posities wél kennen. Andere onderzoekers hebben met onze schattingen aangetoond dat veel fondsen zich beter voordoen dan dat ze zijn met betrekking tot duurzame investeringen (Parise en Rubin, 2025). Onze hoop is dat onderzoek zoals dit indirect leidt tot meer transparantie en verantwoordelijkheidsbesef bij fondsbeheerders.

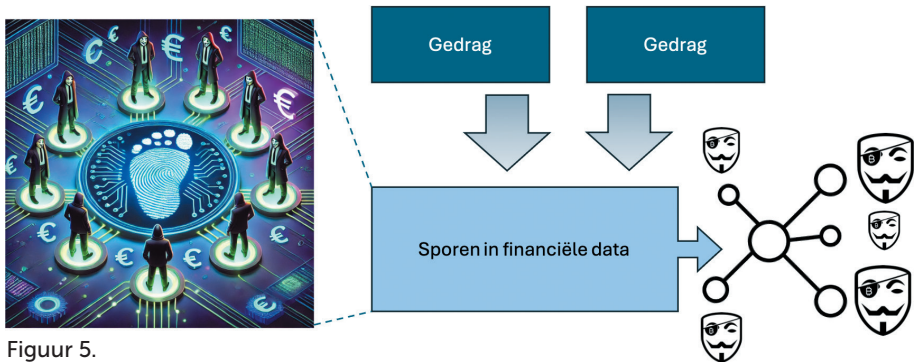
Gezien het belang van financiële administraties kan men zich voorstellen dat de manier waarop deze administraties worden bijgehouden cruciaal is. Idealiter zijn deze systemen gebruiksvriendelijk, schaalbaar én veilig. De af en toe optredende storingen in betaalsystemen laten de problemen zien wanneer zo'n systeem faalt.

De meeste professionele financiële databases zijn tegenwoordig gestructureerde, centrale SQL-databases. Meestal is één partij verantwoordelijk voor het beheer, bijvoorbeeld een bank, verzekeraar of de overheid (Figuur 4). In 2009 kwam echter een nieuwe technologie op: de blockchain. Deze technologie beheert databases decentraal en op een veilige manier. Blockchains kregen brede aandacht via onder andere de succesvolle toepassing ervan in de cryptomunt Bitcoin.



Figuur 4.

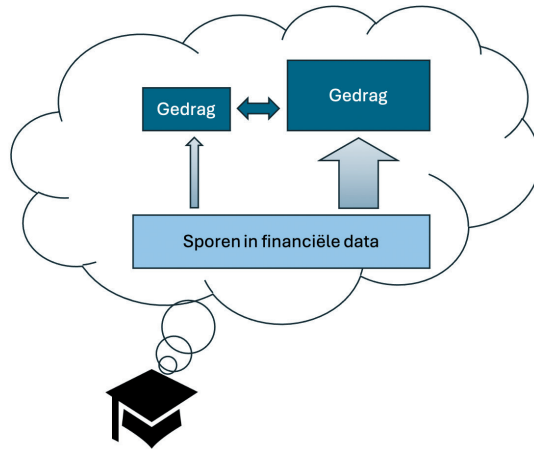
Waar een centrale database vertrouwen vereist in de beheerder, is dit bij een blockchain niet nodig. De data worden redundant opgeslagen en gecontroleerd door meerdere gebruikers met behulp van cryptografie (Figuur 5). Zolang de meerderheid van de gebruikers eerlijk blijft of niet samenwerkt om anderen te benadelen, zijn de data veilig. Sommige blockchains kunnen zelfs programmeercode uitvoeren, waardoor ze als een soort wereldsupercomputer functioneren.



Figuur 5.

Zo'n blockchain kan ook worden gebruikt voor andere waarde-objecten dan crypto, zoals stemrechten. Een voorbeeld hiervan zijn DAOs – digitale organisaties die enkel bestaan op de blockchain. Het besluitvormingsproces van zo'n DAO verloopt via stemmingen van de gebruikers. Samen met Dan Liebau, Thomas Lambert en Peter Roosenboom onderzoeken we of dit stemproces democratischer is dan het stemproces bij traditionele aandeelhoudersvergaderingen, waarin institutionele beleggers doorgaans domineren en kleine aandeelhouders zelden stemmen. Onze resultaten laten zien dat veel gebruikers decentralisatie waarderen, maar dat grote financiële partijen decentralisatie soms slechts ogenschijnlijk nastreven en via stromannen alsnog de DAO domineren.

Toegang tot data, betere verwerking daarvan of superieure infrastructuur levert informatie- en technologievoordelen op die ongelijk verdeeld kunnen zijn in een samenleving. Het effect hiervan op economische uitkomsten kan zowel empirisch als theoretisch worden bestudeerd.



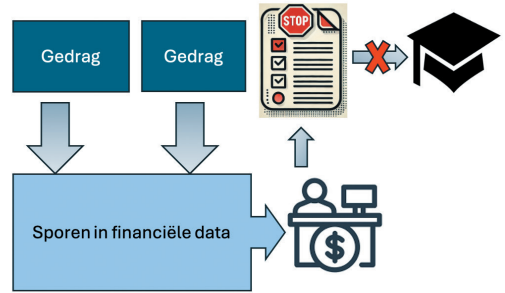
Figuur 6.

Een artikel waarin dit theoretisch wordt behandeld, is een recente publicatie met Mark van Achter in de *Journal of Financial Economics*. We ontwikkelen een theoretisch model van een effectenmarkt waarin technologisch geavanceerde partijen concurreren met minder geavanceerde (Figuur 6). Deze markten werken met nanoseconde nauwkeurigheid en flitshandelaren (ook wel high-frequency traders, HFTs genoemd) spelen hier een grote rol. Deze partijen gebruiken geavanceerde algoritmen om te voorspellen of hun tegenpartij beter geïnformeerd is. HFTs kunnen daardoor vaak verlieslatende transacties vermijden. Hun technologie is echter duur. In onze studie onderzoeken we of de beschikbaarheid van deze technologie goed is voor de markt en samenleving. We vinden dat deze partijen markten vrij snel gaan domineren, zelfs als hun technologie inefficiënt is. Bovendien kunnen ze synchroon gaan reageren op marktsignalen, wat kan leiden tot marktinstabiliteit. Kortom, technologische innovaties kunnen markten en de samenleving verslechteren in plaats van verbeteren.

Een veel gestelde vraag is hoe ik dit vakgebied zie evolueren. Een terugkerende beperking is het gebrek aan hoogwaardige data voor het bestuderen van bepaalde onderzoeksvragen. De data bestaan vaak wel, maar eigenaars mogen of willen deze niet delen met academici, vaak vanwege (terechte) privacyoverwegingen (Figuur 7). Daarom werk ik samen met een TU-Delft-startup, BlueGen, aan een project over synthetische data in finance.



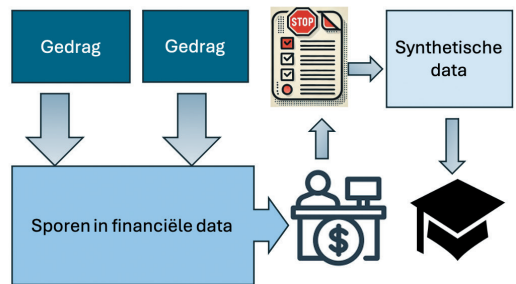
Figuur 7.



Synthetische data zijn afgeleide data met dezelfde statistische eigenschappen als het origineel, maar waarbij herleiding naar de bron (vrijwel) onmogelijk is. Dit kan externe partijen overtuigen om data te delen. Toch is het genereren van dergelijke data niet eenvoudig, zeker niet als specifieke eigenschappen van financiële datasets in ogenschouw worden genomen. In dit project proberen we de beperkingen van bestaande algoritmes die synthetische data genereren te begrijpen en deze algoritmes vervolgens te verbeteren. Zo hopen we uiteindelijk dat studenten en onderzoekers vaker toegang krijgen tot data (Figuur 8).



Figuur 8.

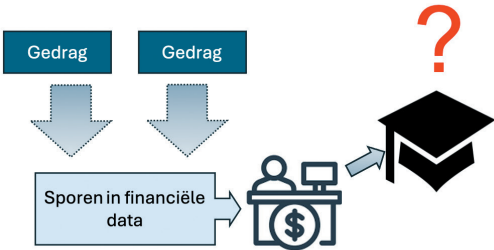


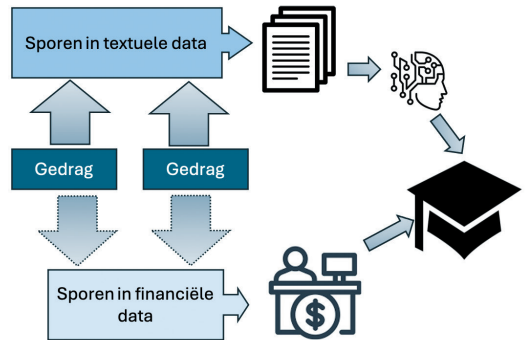
De software voor synthetische data is een vorm van generatieve AI. Natuurlijk kan ik mijn oratie niet voorbijgaan aan de recente AI-modellen, zoals ChatGPT. Eerder beschreef ik het financiële systeem als een administratief systeem. We kunnen specifieke data analyseren om bijvoorbeeld gedrag te reconstrueren. Maar vaak is er ook ongestructureerde tekst beschikbaar. AI-modellen zijn revolutionair omdat ze de betekenis van tekst leren via patronen in de context zonder menselijke interpretatie. Het trainen van deze modellen vereist grote hoeveelheden tekst, maar met het immer uitdijende internet is dat steeds minder een probleem.

Op de universiteit was de introductie van LLMs merkbaar. Studenten waren vroege gebruikers van ChatGPT voor opdrachten. Ook voor ons als onderzoekers is het een waardevol instrument. Zo werk ik met Erik Peek, Stephan Kramer en student Pepijn de Kruijf aan een project over beursanalisten en hun rol in het disciplineren van management. We onderzoeken of kritische analisten beter functioneren. Kritiek is lastig te meten, maar met een taalmodel kunnen we de mate van kritisch denken uit tekstuele en mondelinge communicatie afleiden (Figuren 9 en 10). Dit project loopt nog, maar we hopen deze methode breder toe te passen.



Figuur 9.

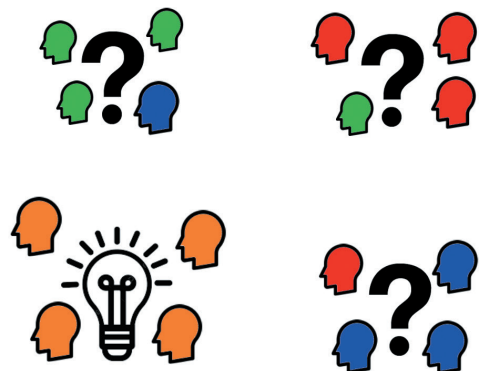




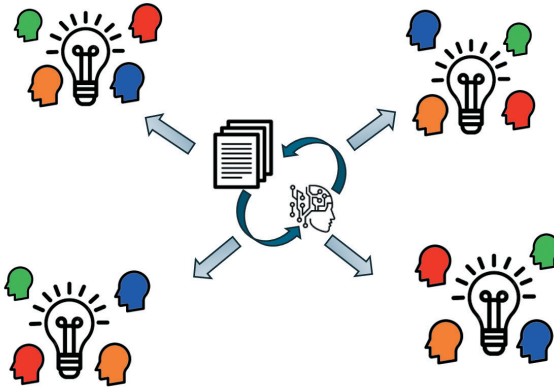
Figuur 10.

De mogelijkheid van algoritmes om ongestructureerde tekst te verwerken is ook belangrijk buiten onderzoek. Een relevant voorbeeld is diversiteit binnen de universiteit en business school. Studies hebben laten zien dat diversiteit leidt tot betere besluitvorming, o.a. door divers denken. Maar “diverse denken” is lastig meetbaar. We vallen dan vaak terug op eenvoudige diversiteitskenmerken zoals etniciteit of geslacht, wat niet altijd effectief is en soms politiek wordt misbruikt.

Tijdens mijn onderwijs merkte ik bijvoorbeeld dat studententeams vaak uit ongelijkwaardige groepen bestonden. Talent klonterde samen, en andere groepen klaagden over moeilijke opdrachten. Studenten leerden ook weinig van elkaar tijdens het groepswork (Figuur 11). De standaardaanpak voor diversiteit met vinkjes voor etniciteit en geslacht faalde hier. Daarom liet ik studenten een profiel schrijven, liet een algoritme clusters maken van profielen die op elkaar leken, en stelde tot slot teams samen met studenten uit de verschillende clusters. Klachten daalden drastisch, en studenten leerden meer (Figuur 12).



Figuur 11.



Figuur 12.

Het zijn deze uitdagingen in onderzoek en de ruimte voor innovatie in onderwijs die mijn werk zo interessant en bevredigend maken. Ik ben dankbaar dat ik dit werk mag doen. Zoals gezegd, ben ik ook dankbaar voor jullie aanwezigheid hier vandaag. Straks geef ik jullie een blijk van waardering in de vorm van een cryptomunt die herinnert aan uw aanwezigheid hier. Scan de QR-code en voeg hem toe aan uw digitale wallet. Het is een administratieve token op de blockchain – geen fysieke munt in uw broekzak.

Deze blijk van waardering staat symbool voor mijn dank aan iedereen die mijn persoonlijke en academische groei mogelijk hebben gemaakt. Ik heb in inspirerende omgevingen gewerkt bij ABN-AMRO, de UvA, de EUR, tijdens diverse research visits, en samen met verschillende partners uit het bedrijfsleven en overheid. Velen hebben mij gesteund en sommigen zijn goede vrienden geworden. Het doet me goed jullie hier te zien.

Naast een stimulerende werkomgeving heb ik ook een groep goede vrienden met wie ik lief en leed deel, en gesprekken voer over de wereld, politiek, familie, liefde, sport, eten en wijn.

En natuurlijk wil ik ook de mensen bedanken die het meest speciaal voor me zijn. Mijn lieve zus, Ilke en zwager Erik-Jan, mijn fantastische ouders en in het bijzonder mijn moeder die een telepathisch vermogen lijkt te hebben om mij te begrijpen en natuurlijk mijn lieve vrouw Jing en geweldige dochter Niki.

Ik heb gezegd.

References/Referenties

Bongaerts, D., Lambert, T., Liebau, D., & Roosenboom, P. (2025). Vote Delegation in DeFi Governance. *arXiv preprint arXiv:2503.11940*.

Bongaerts, D., Roll, R., Rösch, D., van Dijk, M., & Yuferova, D. (2022). How do shocks arise and spread across stock markets? A microstructure perspective. *Management Science*, 68(4), 3071–3089.

Bongaerts, D., & Van Achter, M. (2021). Competition among liquidity providers with access to high-frequency trading technology. *Journal of Financial Economics*, 140(1), 220–249.

Bongaerts, D., van Brakel, J. P., & van Dijk, M. A. (2023). Imputing Daily Mutual Fund Trades. *Available at SSRN 4678139*.

Parise, G., & Rubin, M. (2025). Green window dressing. *The Journal of Finance*, *Forthcoming*.

Dion Bongaerts is a Professor of Financial Technology and Data Analytics at RSM, Erasmus University and Academic Director Fintech at the Erasmus Center for Data Analytics. He specializes in FinTech (with a focus on Blockchain and AI), the behavior of credit rating agencies, the pricing of credit risky instruments, the origins and effects of market illiquidity, and econometrics. His work has been presented at major conferences around the world, including the AFA, WFA, EFA, and NBER meetings and published in top tier academic journals including the Journal of Finance, Review of Financial Studies, the Journal of Financial Economics, Review of Finance, and Management Science. He has received several grants, including a Veni and a Blockchain Research grant from the Dutch National Science Foundation (NWO) and a Lamfalussy Fellowship from the ECB. Prof. Bongaerts holds a PhD degree in Finance from the University of Amsterdam, an MSc in Econometrics from Maastricht University, and has been a visiting scholar at Yale School of Management, University of Pittsburgh, and the Central University of Finance and Economics in Beijing. Moreover, he has several years of professional experience as a risk management quant at ABN-AMRO bank.

In his inaugural address, he discusses the strong interconnection between technology and the financial sector. The whole discussion is centered around data. He provides examples of how such data can be used, explores new ways of managing and organizing such data and connects to novel technological developments such as blockchains, the generation of privacy-friendly synthetic data, and the use of AI, for example to address diversity problems in an atypical context.

ERiM

The Erasmus Research Institute of Management (ERIM) is the joint research institute of Rotterdam School of Management, Erasmus University (RSM) and the Erasmus School of Economics (ESE). The research undertaken by ERIM is focused on management of the firm in its environment, its intra- and interfirm relations, business processes, strategies, finances, consumers, markets and their interdependent connections.

Since its founding in 1999, the objective of ERIM is to carry out first-rate research in management and to offer an advanced doctoral programme in Business and Management. Within ERIM, over 300 senior researchers and PhD candidates are active in five research programmes, spanning all areas of management research. From a variety of academic backgrounds and expertise, the ERIM Community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERiM

ERIM Inaugural Address Series Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl