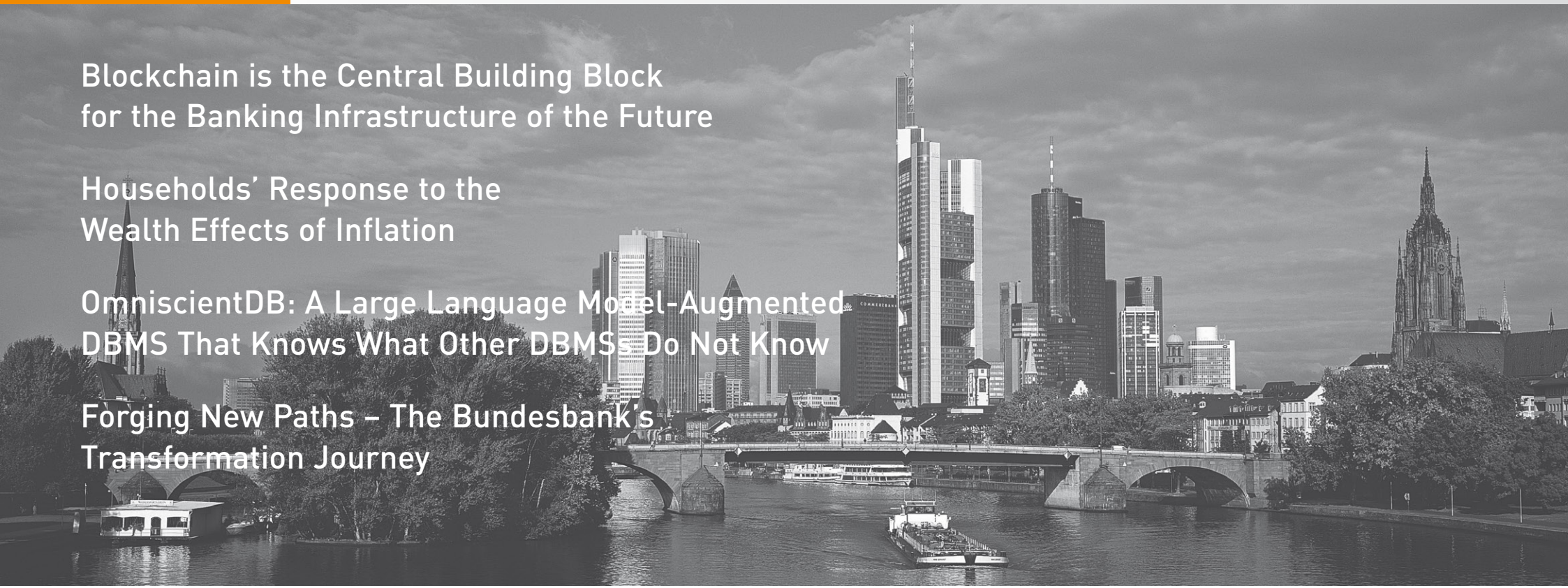
efl insights 02|2023
An efl – the Data Science Institute Publication

efl the Data Science Institute



Blockchain is the Central Building Block for the Banking Infrastructure of the Future

Households' Response to the Wealth Effects of Inflation

OmniscientDB: A Large Language Model-Augmented DBMS That Knows What Other DBMSs Do Not Know

Forging New Paths – The Bundesbank's Transformation Journey

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
www.eflab.de

Gestaltung
Novensis Communication GmbH
Weilburg

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

Printed in Germany
ISSN 1866-1238

Editorial

Blockchain is the Central Building Block for the Banking Infrastructure of the Future

Thomas Ullrich

Thomas Ullrich
Member of the Board of Managing Directors
DZ BANK AG

Blockchain technology has become an indispensable part of the financial world. It promises greater efficiency, speed and security and will substantially change the infrastructure of banks.

In recent years, DZ BANK has continuously invested in the further development of processes and services to strengthen its Blockchain expertise. Although the macroeconomic environment is becoming gloomier with factors such as high inflation and the impact of higher interest rates, we persist in our work on forward-looking projects. Blockchain technology plays a crucial role in this, as it is a central building block for the future banking infrastructure. In recent years, DZ BANK has identified an increasing number of applications for Blockchain technology, particularly in the capital markets business and payments.

Payment transactions in particular have undergone major changes in recent years, including instant payments, crypto-assets and the prospect of the Digital Euro. However, these developments bring with them challenges such as increasing cost pressure and high efficiency requirements. Blockchain technology enables us to simplify processes and create new services for our customers.

An example of this is the Commercial Bank Money Token (CBMT), which modernises traditional commercial bank money and creates a stable and secure form of Blockchain-based tokenised money. The need for Blockchain-based fiat money is great in both commerce and industry. DZ BANK has already developed a trigger solution that acts as an interface between Blockchain technology in companies and traditional payments. Although not all companies are sufficiently prepared for these changes yet, a positive trend is emerging, especially among technically savvy companies such as those in the mechanical engineering sector.

The European Central Bank (ECB) plans to introduce a Digital Euro threatening the two-tier banking system. Currently, there are considerations to design the Digital Euro as a wallet based on instant payment – also as a card that can be refilled at any time. This could lead to an outflow of bank deposits and the loss of an important source of refinancing. In order to preserve the stability of the financial system, it is important to maintain the functional separation between the ECB and commercial banks.

In addition to payment transactions, Blockchain technology also has enormous benefits for capital markets business, as transactions become more efficient, cheaper and more secure thanks to the decentralised database. With increasing digitisation of securities, the need for crypto custody is also growing. At DZ BANK, we will launch a solution for direct investment in crypto-assets in the course of the year. We are also implementing a crypto custody platform for institutional clients and the bank itself and will thus handle and hold digital financial instruments of all kinds. In mid-February, we invested in a Blockchain-based crypto security of Siemens AG for the first time ourselves. We will stick to this offer despite the recent turmoil on the crypto markets. After all, demand is high among our customers. In this context, it is even more important that our customers are aware of the risks of the investment. That is precisely why we provide them with targeted information.

Developments with regard to Blockchain will continue to accelerate. While the high volatility of Bitcoin and other crypto-assets reliably produce headlines and controversy, more and more real-world applications will be found across a range of industries in the future. It is in the financial sector's own interest to take the lead here and test and exploit the potential of Blockchain. After all, nothing less than the future of banking infrastructure is at stake.

Research Report

Households' Response to the Wealth Effects of Inflation

WE STUDY REDISTRIBUTIVE EFFECTS OF INFLATION USING A RANDOMIZED INFORMATION EXPERIMENT ON BANK CLIENTS. ON AVERAGE, INDIVIDUALS ARE WELL INFORMED ABOUT CURRENT INFLATION AND ARE CONCERNED ABOUT ITS IMPACT ON WEALTH. YET, MOST INDIVIDUALS ARE NOT AWARE OF HOW INFLATION ERODES NOMINAL POSITIONS. ONCE THEY RECEIVE INFORMATION ON THIS EROSION CHANNEL, THEY UPDATE PERCEPTIONS AND EXPECTATIONS ABOUT OWN NET NOMINAL POSITIONS. LEARNING ABOUT THE INFLATION-INDUCED EROSION OF NOMINAL POSITIONS CAUSALLY AFFECTS CHOICES IN HYPOTHETICAL REAL-ESTATE TRANSACTIONS AND ACTUAL CONSUMPTION. THE FINDINGS SUGGEST THAT HOUSEHOLD WEALTH MEDIATES THE SENSITIVITY OF CONSUMPTION TO INFLATION ONCE HOUSEHOLDS ARE AWARE OF THE BALANCE-SHEET EFFECTS OF INFLATION.

Philip Schnorpfeil
Michael Weber
Andreas Hackethal

Introduction

After having been dormant for decades, inflation reached levels in 2022 not seen by many households during their lifetimes. On average, households dislike high levels of inflation because nominal incomes usually do not keep up with sharp increases in inflation in the short run. However, unexpected increases in the price level do not only generate losers but can also benefit households, among other economic actors. Surprise inflation lowers the real value of nominal assets but also erodes the real value of debt positions with fixed nominal interest obligations (Doepke and Schneider, 2006), redistributing wealth from savers to borrowers. We label this mechanism as erosion channel of nominal positions. So far, little is known about whether households are aware of these distributional consequences of inflation and how they adjust their economic decisions to news about their real net worth.

We study the extent to which households are aware of the erosion channel of inflation, how they adjust perception and expectation of their own net nominal position, as well as how news about their real wealth affect housing and leverage choices in a hypothetical real-estate investment task and their planned and actual consumption.

Methodology

To study these questions, we run a large, randomized control trial (RCT). We implemented the RCT on several thousand bank clients in July 2022 when inflation in Germany was at 8.7%, a 70-year high. The survey respondents are on average better educated than the average German in the population, they have large nominal positions (e.g., 55% have outstanding debt), accurate perceptions of current inflation, and they are concerned about its impact on their wealth. In the survey experiment, we first ask respondents questions on their pre-existing knowledge, such as how various balance-sheet items fare following an unexpected increase in inflation. Respondents also estimate their recent change in real net worth and decompose their balance sheet so we can calculate their net nominal position.

The subsequent information-provision experiment builds the core of our survey. We randomly assign respondents to three groups: two treatment groups and one control group. The treatment groups receive (i) information on the current inflation rate, (ii) an explanation that higher inflation hurts savers / benefits debtors because it erodes nominal positions, and (iii) a calculation of the current change in the real value of representative savings products or loans. The single difference between the two treatments is that one discusses the erosion of savings, while the other is on loan erosion. The control group only receives information about (i), current inflation. Hence, we inform all survey participants about prevailing inflation rates but only subjects in the treatment conditions learn about the nominal erosion channel of inflation. By comparing outcomes across treatment and control groups, we can thus isolate the effect of information about the erosion channel of inflation. Post-treatment, we elicit beliefs about nominal positions, own real net wealth, and the economy. Moreover, we ask respondents about spending plans and hypothetical real-estate investments, and we can track actual consumption choices using administrative data provided by our bank partner.

Finding 1: Existing Knowledge

One-quarter of respondents believe the impact

of surprise inflation on fixed-interest savings products is very negative, and one-half believe it is negative (see Figure 1). This indicates some prior awareness about inflation-induced savings erosion, though the beliefs are close to those about stocks, which individuals believe to provide some protection against inflation (Schnorpfeil et al., 2023). Knowledge about loan erosion is more limited. Only 9% of respondents believe the impact is very positive, and 25% believe it is positive. These findings indicate limited scope for aggregate consequences of the inflation-induced debt erosion, as households are unlikely to act on effects they are largely unaware of.

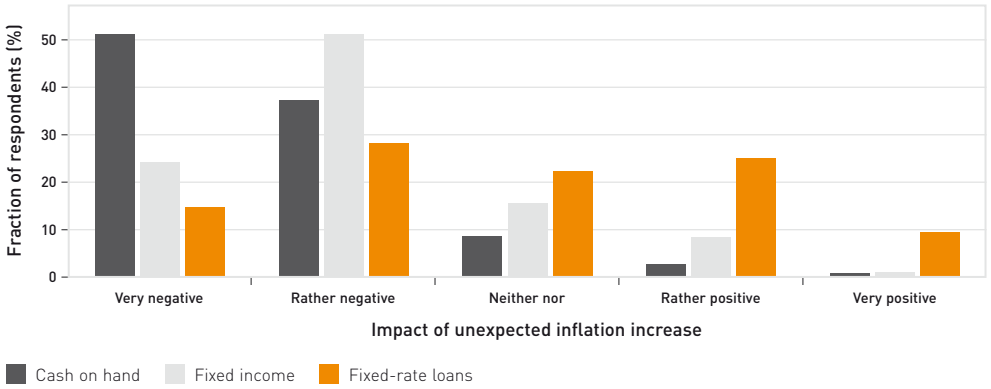


Figure 1: Prior Knowledge about the Wealth Effects of Inflation

Finding 2: Learning Affects Real-net-wealth Perception

Randomized provision of information on the erosion channel of nominal positions affects estimates of own real wealth and beliefs about these nominal positions. Conditioning on pre-treatment estimates of real-wealth changes, we find respondents who receive the savings-erosion treatment on average estimate their change in real wealth weakly lower than respondents in the control group. Respondents learning about inflation-induced loan erosion increase their estimate of the real-wealth change relative to the control group by three percentage points. Those respondents with larger exposure to the nominal position they learn about drive these effects: individuals with a negative nominal position ("net debtors") react more strongly to the loan treatment. For example, among net debtors, those who receive the loan treatment estimate their real-wealth change five percentage points more positively on average than net debtors in the control group.

Finding 3: Effects on Consumption

We show that learning about the wealth effect of inflation affects planned and actual consumption. Changes in real-wealth estimates, caused by the loan-erosion treatment, significantly impact planned and actual discretionary spending. We arrive at this result using an instrumental-variables approach, exploiting the treatment as a source of exogenous variation in perceived wealth. This finding implies that real-wealth changes affect households' consumption response to inflation, conditional on households being aware of balance-sheet effects of inflation.

Conclusion

We causally study the inflation-induced erosion of nominal wealth using a survey-based RCT. On average, individuals are highly concerned about the impact of inflation on their real net worth, and they are well informed about prevailing inflation levels. Yet, they know surprisingly little about the reduction in the real value of nominal positions due to surprise inflation. Once we inform individuals about the erosion channel of nominal positions, they update their current and expected real net worth and causally change their consumption and leverage choices in a hypothetical real-estate transaction. Our results indicate the redistributive nature of inflation across households and provide causal estimates for how individuals adjust behavior following inflation-induced redistribution.

References

Doepke, M.; Schneider, M.:
Inflation and the Redistribution of Nominal Wealth.
In: Journal of Political Economy, 114 (2006) 6, pp. 1069–1097.

Schnorpfeil, P.; Weber, M.; Hackethal, A.:
Inflation and Portfolio Choices.
Working Paper, 2023.

Research Report

OmniscientDB: A Large Language Model-Augmented DBMS That Knows What Other DBMSs Do Not Know

WE PRESENT OUR VISION OF OMNISCIENTDB, A NOVEL DATABASE THAT LEVERAGES THE IMPLICITLY STORED KNOWLEDGE IN LARGE LANGUAGE MODELS TO AUGMENT DATA SETS FOR ANALYTICAL QUERIES OR MACHINE LEARNING TASKS. OMNISCIENTDB EMPOWERS USERS TO AUGMENT DATA SETS BY MEANS OF SIMPLE SQL QUERIES AND THUS HAS THE POTENTIAL TO DRAMATICALLY REDUCE THE MANUAL OVERHEAD ASSOCIATED WITH DATA INTEGRATION. IT USES AUTOMATIC PROMPT ENGINEERING TO CONSTRUCT APPROPRIATE PROMPTS FOR GIVEN SQL QUERIES AND PASSES THEM TO A LARGE LANGUAGE MODEL LIKE GPT-3 TO CONTRIBUTE ADDITIONAL DATA, AUGMENTING THE EXPLICITLY STORED DATA. OUR INITIAL EVALUATION DEMONSTRATES THE GENERAL FEASIBILITY OF OUR VISION, EXPLORES DIFFERENT PROMPTING TECHNIQUES IN GREATER DETAIL, AND POINTS TOWARDS FUTURE RESEARCH.

Matthias Urban

Duc Dat Nguyen

Carsten Binnig

Introduction

Traditionally, relational databases are required to explicitly capture the reality of a particular domain, meaning that all relevant facts need to be stored in the database in order to be queried by the user. However, this so-called closed-world assumption significantly limits the ways in which the information in a database can be used. For example, think of a database that stores information about movies and their revenues. While a breakdown of the revenue by director might be an interesting query a user wants to issue, this information might not be stored in the database. Today, the only way to make additional information available for querying or other downstream tasks (e.g., training a Machine Learning (ML) model) is to explicitly integrate additional data sources into the database, which requires extensive manual efforts for every data source.

Idea and Simple Example

In this paper, we thus present our vision of OmniscientDB, an open-world Data Base Management System (DBMS) that can automatically augment existing databases with world knowledge for the execution of Structured Query Language (SQL) queries. To do so, OmniscientDB can not only generate additional tables on-the-fly but also complete existing tables with user-requested rows or columns. To enable this, OmniscientDB makes use of the world knowledge that is implicitly stored in large language models (LLMs, e.g. GPT-3 (Brown et al., 2020)).

To illustrate the benefits of OmniscientDB by an example, imagine a data scientist trying to analyze the revenues of recent movies as mentioned before. For the questions of the data scientist, important information like the starring actors or directors, however, is not contained in the data set, despite their potentially large impact on the movies' revenues. While traditionally such information would need to be explicitly integrated first, with OmniscientDB the data scientist could simply use the knowledge stored in the LLM for augmentation. For instance, OmniscientDB could automatically generate the missing information about the actors starring in the movies, allowing a more extensive analysis.

Virtual Tables

OmniscientDB leverages the knowledge implicitly stored in the parameters of LLMs for augmentation and makes it available for querying via SQL using so-called virtual tables. Virtual tables can be treated by users just like traditional tables in relational databases, e.g. they can be used as operands in traditional query operators.

However, they are not explicitly stored in the database but instead act as a proxy for the knowledge stored in LLMs. For instance, in the example above, the user is able to join the movies table with the virtual actors table to gen-

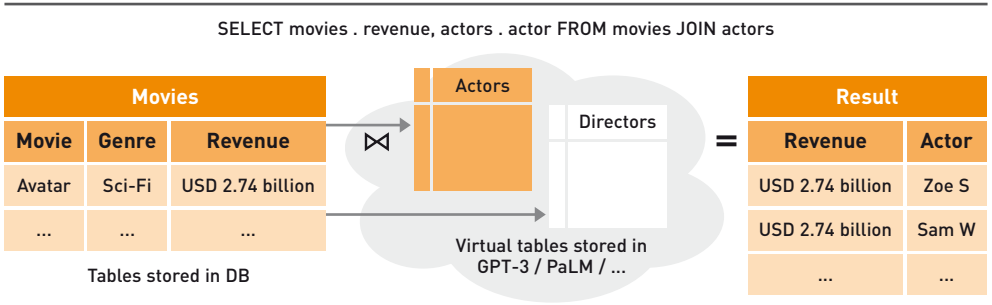


Figure 1: SQL Query That Joins a Movies Table Stored in the Database with an Actors Table Implicitly Stored in the Weights of a LLM (e.g., GPT-3).

erate additional information about actors, as shown in Figure 1. The information about actors is materialized on-the-fly during the execution of the query.

Virtual tables are an easy way for the data scientist to also take actor information into account to perform the analysis.

In particular, it is important to mention that this way of augmenting data sets comes with almost zero manual overhead. The data scientist just has to formulate a SQL query involving virtual tables, and the database will automatically generate the necessary information to perform the join.

The knowledge available in OmniscientDB via virtual tables is obviously bound by the knowledge stored in LLMs. However, recent LLMs have been trained on huge data sets and by now scale to hundreds of billions of parameters (Brown et al., 2020; Chowdhery et al., 2023), leading to tremendous amounts of knowledge stored in their parameters. In the Challenges Section, we will discuss further opportunities how LLMs can tap into knowledge not stored in their parameters.

Case Study

To show the general feasibility of OmniscientDB, we present a case study in which we examine the quality of materialized LLM knowledge for virtual tables. To materialize knowledge from LLMs, we automatically generate so-called prompts, which are short strings we provide as input to the LLM. The LLM is trained to complete such prompts and thereby reveal the knowledge stored inside of it. In the case study, we examine different variations on how prompts for virtual tables can be generated on different real-world data sets. All experiments for the case study are performed with the GPT-3 davinci model by OpenAI.

Structured Prompts vs. NL Prompts

Our default way of creating prompts to extract data for columns of a virtual table is to linearize entire table rows to obtain a (structured) prompt as follows:

Title: Ant man | Year: 2017 | ... | Actors:

This prompt would let the LLM generate all main actors appearing in Ant man. However, since LLMs are pre-trained on natural language, we might obtain better results by using prompts that are the beginning of natural language sentences (NL prompts). Prompts in NL for database operations are hard to obtain though since they require formulating the information in the table as NL sentences. They either have to be created manually, or by using templates, or they could be the output of a sequence-to-sequence ML model. On the other hand, they might result in generated values of higher quality.

To find out if it makes sense to put the effort into constructing good NL prompts, we compare structured prompts with NL prompts in a simplified setting. We consider a table with a single explicitly stored column and would like to add more columns using LLM knowledge. For data sets, we use the cities data set (1,000 largest cities; Kaggle, 2017), where the single column is the *City Name* and we want to add columns for *Longitude* and *Latitude*.

Additionally, we use the IMDB Movies data set (first 1,000 movies; OpenML, 2022), where the single column is the movie *Title*, and the LLM should generate values for the *Release Year*, *Genre*, *Runtime*, *Actors* and *Director*. Hence, we cover two wildly different domains and a variety of data formats such as text, float numbers, and dates. In this case, the structured prompts, as introduced before, result in prompts like *Title: Avatar | Runtime:*, while the NL prompt is *The runtime of Avatar is*. For this initial evaluation, we measure how many cells are filled correctly and report an accuracy value averaged over all cells. Longitude and Latitude (for the city data set) are considered correct if the difference to the ground truth value is below 0.1 and for columns where we expect multiple values (e.g., Genre) cells are considered correctly filled if the model predicts at least one correct value.

The result (Table 1) shows that, perhaps surprisingly, structured prompts perform better than NL prompts. For the movies table the difference is almost 20% in accuracy. Hence, we decide to use structured prompts, since they are easier to generate and yield more accurate generated values.

	Cities	Movies
Natural Language Prompt	64.65%	51.70%
Structured Prompt	71.65%	71.58%

Table 1: Accuracy of Natural Language Prompts and Structured Prompts on the Two Data Sets Cities and Movies.

Variants of Structured Prompts

Next, we investigate what information should be part of the prompt to obtain generated values of high quality. We fix the type of prompt to structured prompts and consider two kinds of information to add to the prompt: row information and example values. In our prompt, as described in the beginning of this section, we linearize entire rows to obtain the prompts and thus already include the full row information in the prompt (Prompt: *Title: Ant man | Year: 2017 | ... | Runtime:*).

We compare this prompt with a minimal prompt that only includes the value for the first attribute (Prompt: *Title: Ant man | Runtime:*). On top of that, we explore how additional example values (e.g., runtimes of other movies) affect the quality of the extraction. Hence, we construct prompts that always begin with two example rows that contain a value for the column to be materialized. For the movies data set, a prompt might look like this:

Title: Avengers: Endgame | ... | Runtime: 2h 23m
Title: Spiderman: Homecoming | ... | Runtime: 2h 13m
Title: Avatar | Runtime:

Table 2 shows that prompts including row information help the model identify the piece of knowledge that it should generate and provide example values whenever they are available.

	Cities	Movies
minimal	71.65%	71.58%
+ row info	75.00%	80.18%
+ examples	73.65%	79.78%

Table 2: Accuracy of Different Structured Prompts on the Two Data Sets Cities and Movies (+ Row Info Includes Additional Information from the Same Row; + Examples Includes Example Values for the Column to Be Materialized).

Challenges

Due to the vast amount of knowledge stored in the parameters of LLMs, they present ample opportunities to augment existing data sets as presented in our case study. However, they also pose unique challenges as we explore in the following:

External Knowledge

While large LLMs store large amounts of publicly available knowledge, they are still far from being omniscient. In particular, knowledge has to be present sufficiently often in the pre-training data set such that the LLM is able to recall it. To make OmniscientDB live up to its name, it has to be able to also tap into external knowledge not stored in its weights. For instance, if LLMs are able to access open data sets on the Web or private data sets in data lakes, they would be able to augment the data sets at hand much more effectively. Fortunately, work on retrieval-augmented language models (Guu et al., 2020) has shown that it is in principle possible to let LLMs utilize external knowledge. More recently, researchers have even proposed language models that search the Web (Nakano et al., 2023). However, it is not yet clear how LLM can be used to retrieve structured data sets (i.e., tables) and not only NL passages from external corpora.

Trust & Hallucination

Probably the most critical challenge is that LLMs are known to suffer from hallucination, a phenomenon where LLMs generate non-factual statements. In the context of OmniscientDB, this means that it might generate values during the materialization of virtual tables that sound plausible but are not actually correct. This not only reduces users' trust in the augmentation of OmniscientDB but might also negatively affect downstream analytical queries or ML applications. However, it is already possible to extract well-calibrated certainty scores from LLMs (Kadavath et al., 2023), which could be used to detect hallucinations. As such, there already exist techniques to mitigate the risks of LLMs and future research will further reduce the amount of generated values that are factually wrong. For applying these trends to the generation of structured data like table rows or columns, as considered here, again additional research will be needed.

Conclusion

OmniscientDB is our vision of how the knowledge stored in the parameters of LLMs can be integrated into databases. The concept of virtual tables allows users to seamlessly integrate such knowledge into their existing data sets with just a few SQL queries and without any manual overhead. In our case study, we showed the general feasibility of our ideas by experimenting with different automatically-generated prompts to materialize virtual tables and columns. However, several interesting research challenges lie ahead to further enhance the usefulness of OmniscientDB. In particular, we are interested in how language models that are able to independently search the Web could enable an entirely automated data augmentation process, where materialized columns and tables are not bound by the knowledge that can be stored in model parameters, but by knowledge accessible via the Internet.

References

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; et al.:
Language Models Are Few-Shot Learners.
In: Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS); Vancouver, Canada, 2020.

Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; et al.:
Palm: Scaling Language Modeling with Pathways.
Working Paper, 2023.

Guu, K.; Lee, K.; Tung, Z.; Pasupat, P.; Chang, M.:
Retrieval Augmented Language Model Pre-training.
In: Proceedings of the 37th International Conference on Machine Learning (ICML); Vienna, Austria, 2020.

Kadavath, S.; Conerly, T.; Askell, A.; Henighan, T.; Drain, D.; et al.:
Language Models (Mostly) Know What They Know.
Working Paper, 2023.

Kaggle:
World Cities Database, https://www.kaggle.com/datasets/max-mind/world-cities-database?resource=download, 2017.

Nakano, R.; Hilton, J.; Balaji, S.; Wu, J.; Ouyang, L.; et al.:
Webgpt: Browser-Assisted Question-Answering with Human Feedback.
Working Paper, 2023.

OpenML:
IMDB Movies Dataset, https://www.openml.org/search?type=data&status=active&id=43603&sort=runs, 2022.

Insideview

Forging New Paths – The Bundesbank's Transformation Journey

INTERVIEW WITH KARMELA HOLTGREVE

Karmela Holtgreve
Director General Strategy and Innovation
Deutsche Bundesbank

The Deutsche Bundesbank joined the efl in May 2023. We would like to take this opportunity to extend a warm welcome to your institution. Ms. Holtgreve, you will represent the Bundesbank on the efl's Board of Directors. What do you hope to gain from your membership?

As an institutional partner, we hope to benefit from the skills and expertise present in the efl. We see our membership as an opportunity to further strengthen our own analytical capabilities and foster new partnerships outside the central banking universe. We want to play an active role ourselves in developing cooperative networks and technical solutions.

Beyond that, our colleagues are already looking forward to the training events offered by the efl and to meeting many talented people.

You mentioned fostering networks outside the "central banking universe". Is this something that the Bundesbank is looking to promote more in the future?

Yes, we want to broaden our horizons. That is why, for example, we are stepping up our collaborations with start-ups from the FinTech scene. But it doesn't stop there. The digital revolution presents us with challenges that do not affect central banks alone. For me, it is only logical that we try to solve these challenges together with a diverse range of partners. Innovation can come from within – but sometimes it requires an outside force.

Since the start of 2023 you have been in charge of the newly founded Directorate General Strategy and Innovation. What is your impression, how do the Bundesbank and innovation fit together?

Fortunately, the fit is getting better and better. The Bundesbank stands for stability and tradition. Stability requires constant questioning, which is why we cannot afford to stand still. And the world around us certainly does not. In particular, the world of finance is evolving at a rapid pace. This is also one reason why, at the start of 2023, we launched the largest transformation program the Bundesbank has seen in decades.

As I mentioned earlier, the digital revolution is shaking up business models and the way we work together. We came to the conclusion that if we wish to continue to fulfill our legal mandate in the best possible way, we need to make the Bundesbank future proof.

Let us take a look at the near future. Are there any specific use cases that bring innovation to the Bundesbank and that you are already working on?

To give an example, we have recently started to work on two prototypes that involve the use of ChatGPT and other large language models. Colleagues there have to deal with vast amounts of information and reports. They need fast and accurate access to information from these enormous reports in order to do their work.

We have collected a large number of these publicly available documents and news articles. We use these in combination with large language models in an internal Bundesbank chatbot. This allows us to efficiently identify text passages to find information, answer questions and help the business unit in their work. Our colleagues can then engage with it for efficient information extraction and interactive Q&A, further improving their work.

In general, we try to help our colleagues to use innovations like ChatGPT in a sensible and safe way.

Thank you for this interesting conversation.

Infopool

News

EFL ANNUAL CONFERENCE 2023

The efl Annual Conference on "Opportunities and Challenges of Generative AI" will take place on November 15th, 2023. We cordially invite you to participate. The conference is organized by Prof. Dr. Bernd Skiera and his team: Karlo Lukic, Lennart Kraft, Yuxi Jin, Lukas Jürgensmeier, Maximilian Matthe, and Timo Müller-Tribbensee. For more information, please visit: **https://www.eflab.de/events/efl-annual-conference-2023**.

Deutsche Bundesbank as New efl Industry Partner
We are pleased to announce that Deutsche Bundesbank joined the efl as new industry partner in May 2023. Deutsche Bundesbank is located in Frankfurt/Main, Germany, and as part of the Eurosystem it shares responsibility with the other national central banks and the ECB for monetary policy in the euro area with the aim to secure price stability. In addition, the Bundesbank performs other key tasks at both the national and international level. Among these are national supervision of credit institutions, including a role in the European Single Supervisory Mechanism, as well as cash management, payment systems and financial stability. Welcome to the efl!

M.Sc. Florian Ewald Joined the efl in May 2023
Florian Ewald studied Business Administration at Goethe University and received his Master's degree in early 2023. Subsequently, he joined the Chair of e-Finance (Prof. Dr. Peter Gomber). His research focus is on machine and deep learning applied to high-frequency trading and market microstructure related topics. Welcome to the efl!

Prof. Dr. Kevin Bauer Becomes Junior Professor at University of Mannheim
Former efl team member Kevin Bauer starts his Chair of E-Business and E-Government at the University of Mannheim. His main fields of interests are human-machine interaction, human-centered and applied machine learning as well as explainable AI. The efl wishes him much success and all the best in his new role and for his future!

ZEVEDI Supports Project Led by Prof. Dr. Oliver Hinz
The Centre Responsible Digitality (ZEVEDI) supports the project "Digitalization of Corporate Communications (DigUKom)" led by Prof. Dr. Oliver Hinz and his team. The project's goal is to derive normative recommendations based on empirical findings in the area of corporate communication by means of an extensive literature review.

Emily Kormanyos and Colleagues Received Vernon L. Smith Young Talent Award
Emily Kormanyos, Daniel Worring, and Jakob Famulok (Chair of Prof. Dr. Andreas Hackethal) were awarded the 2023 Vernon L. Smith Young Talent Award by the Society for Experimental Finance for their proposal "Moral Licensing through ESG Investments: Causal evidence on individual preferences and actions".

Selected efl Publications

Bender, M.; Clapham, B.; Schwemmlein, B.:
Shifting Volumes to the Close: Consequences for Price Discovery and Market Quality.
In: 32nd European Financial Management Association Conference (EFMA); Cardiff, UK, 2023.

Costola, M.; Nofer, M.; Hinz, O.; Pelizzon, L.:
Machine Learning Sentiment Analysis, COVID-19 News and Stock Market Reactions.
In: Research in International Business and Finance, 64 (2023), Article No. 101881.

Gnewuch, U.; Morana, S.; Hinz, O.; Kellner, R.; Mädche, A.:
More than a Bot? The Impact of Disclosing Human Involvement on Customer Interactions with Hybrid Service Agents.
Forthcoming in: Information Systems Research.

Gomber, P.; Sagade, S.; Theissen, E.; Weber, M. C.; Westheide, C.:
Spoilt for Choice: Determinants of Market Shares in Fragmented Equity Markets.
In: Journal of Financial Markets, 64 (2023), Article No. 100816.

Hättasch, B.; Bodensohn, J.-M.; Vogel, L.; Urban, M.; Binnig, C.:
WannaDB: Ad-hoc SQL Queries over Text Collections.
In: Proceedings of the 20th Conference on Database Systems for Business, Technology and Web (BTW); Dresden, Germany, 2023.

Jürgensmeier, L.; Skiera, B.:
Measuring Fair Competition on Digital Platforms.
Working Paper, 2023.

Langenecker, S.; Schalles, C.; Sturm, C.; Binnig, C.:
Steered Training Data Generation for Learned Semantic Type Detection.
In: Proceedings of the 2023 International Conference on Management of Data (SIGMOD); Seattle (WA), US, 2023.

Lenz, G.; Mayer, M.:
Hollywood, Wall Street, and Mistrusting Individual Investors.
Forthcoming in: Journal of Economic Behavior and Organization.

Menkveld, A. J.; Dreber, A.; Holzmeister, F.; Clapham, B.; Lausen, J.; et al.:
Non-standard Errors.
Forthcoming in: Journal of Finance.

Meyer, S.; Uhr, C.; Loos, B.; Hackethal, A.:
Switching from Commissions on Mutual Funds to Flat-Fees: How Are Advisory Clients Affected?
Forthcoming in: Journal of Economic Behavior and Organization.

Siering, M.:
Peer-to-Peer (P2P) Lending Risk Management: Assessing Credit Risk on Social Lending Platforms Using Textual Factors.
Forthcoming in: ACM Transactions on Management Information Systems

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

Infopool

RESEARCH PAPER:
NUMEROLOGICAL HEURISTICS AND CREDIT RISK IN PEER-TO-PEER LENDING

This paper investigates whether and how different heuristics have distinct effects in the context of peer-to-peer (P2P) lending. Based on existing theories, the authors hypothesize that the funding success and repayment performance differ depending on the heuristics used by borrowers. These heuristics are determined by the borrowers' distinct motives used to set their loan amounts. Two numerological heuristics, the round-number heuristic and the lucky-numbers heuristic, are examined using data from a Chinese P2P lending platform. The authors find that round-number loans are less likely to get funded and exhibit poor repayment performance once funded, whereas lucky-number loans exhibit the opposite pattern. These results challenge the conventional understanding of heuristics as behavioral biases and provide new insights into their heterogeneity and implications for the credit market.

Hu, M. R.; Li, X.; Shi, Y.; Zhang, X. M.
Forthcoming in: Information Systems Research.

RESEARCH PAPER:
A SENTIMENT MEASURE CONSTRUCTED FROM ISSUANCES OF RETAIL STRUCTURED EQUITY PRODUCTS

This paper constructs a new cross-sectional sentiment measure for large capitalization stocks, using retail structured equity product (SEP) issuances. The SEP sentiment measure predicts negative abnormal returns on the SEP reference stocks based on a variety of factor models, and also predicts returns in Fama-MacBeth regressions that include a wide range of covariates. The results show significant abnormal returns of calendar-time portfolios of between -1.09% and -1.25% per month. Consistent with the notion that SEP issuances reflect investor sentiment, aggregated SEP issuances are highly correlated with the Baker-Wurgler sentiment index. Tobit regressions reveal that proxies for attention and sentiment predict SEP issuance volumes, providing additional evidence consistent with the hypothesis that SEP issuances reflect sentiment.

Henderson, B. J.; Pearson, N. D.; Wang, L.
Forthcoming in: Journal of Finance.

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 (> 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

Further information about the efl is available at 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.

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, DEUTSCHE BUNDESBANK, 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

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

or visit our website
http://www.eflab.de