efl insights 01|2026
An efl – the Data Science Institute Publication

efl | the Data Science Institute

The British Getty Ruling and Its Implications for Europe's AI Regulation

Response of Debtors to Rate Changes

Using LLMs for Data Engineering in Enterprises Is Challenging

Delivering the EU Equities Consolidated Tape

DEUTSCHE BÖRSE GROUP
DZ BANK Gruppe
AMERICAN EXPRESS
DEUTSCHE BUNDESBANK EUROSYSTEM
finanz informatik

GOETHE UNIVERSITÄT FRANKFURT AM MAIN
TECHNISCHE UNIVERSITÄT DARMSTADT

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

1. Ausgabe, 2026
Auflage: 200 Stück (Print) / 2.000 Stück (E-Paper)

Printed in Germany
ISSN 1866-1238

Editorial

The British Getty Ruling and Its Implications for Europe's AI Regulation

Sabina Jeschke & Tobias Fuchs

Prof. Dr. Sabina Jeschke
CEO, KI Park e.V.; Supervisory Board Member of Rheinmetall AG & Aumovio SE

Dr. Tobias Fuchs
General Counsel & Founding Partner THE MAKERY Group; AI-driven Legal, Governance & Compliance

The development of powerful AI models has triggered an intense regulatory debate in recent years. A central point of contention is the relationship between models and the data on which they are trained. Much of Europe's data protection and copyright logic relies on the assumption that models "contain" personal or protected content, at least in abstracted form, and could therefore potentially disclose it. Against this background, a recent judgment by the High Court of England has attracted particular attention.

In *Getty Images vs. Stability AI*, the Court held that generative models do not store or reproduce training data in the conventional sense. Instead, they compress patterns into mathematical parameters that describe relationships without retaining the underlying data itself. This distinction may appear technical, but it constitutes a fundamental challenge to widely held regulatory assumptions. It affects both copyright considerations and the assessment of risks posed by AI models under data protection law.

The ruling thus lies at the intersection of two perspectives: On one side, a European discourse that tends to treat models as abstract databases and focuses on the potential reconstruction of personal information. On the other, the British judgment emphasizes the technical reality of statistical learning processes, which generalize patterns without preserving concrete content. This divergence is more than a legal nuance; it touches on the conceptual foundations of Europe's emerging AI regulatory framework.

If the British perspective prevails, it would enable a clearer separation between data and models, with immediate practical consequences. Under certain conditions, AI training processes might no longer be classified as "further processing of personal data" if such data cannot be reconstructed from the model. This could, for example, make it possible to transfer or repurpose models without repeatedly triggering the full set of data protection obligations tied to the original data. Technical assessment procedures – such as formalized leakage tests, reconstruction analyses or risk evaluations based on model behaviour – could carry greater regulatory weight than formal consent or purpose-limitation mechanisms. For many European AI applications, this would create greater legal certainty: models could be used, retrained or audited more frequently without the training process itself becoming a regulatory barrier. At the same time, real risks to individuals would remain addressable, but on the basis of technical evidence rather than hypothetical assumptions.

For data-centric industries – including financial services, insurance, manufacturing, energy, healthcare and mobility – such clarification would have significant operational implications. Many of these sectors rely on large and heterogeneous data assets whose use for AI training is currently limited by legal uncertainty. A regulatory framework that more accurately reflects how models function could facilitate more frequent internal retraining, cross-sector pattern analysis or model transfer within organizations without repeatedly reopening full data protection procedures. Audit mechanisms could focus more directly on verifiable model characteristics – such as whether a model memorizes information or whether statistical patterns allow meaningful reconstruction. This would not reduce regulatory safeguards but sharpen them: protection would be applied where real risks arise, while innovation-inhibiting redundancies could be avoided.

Europe now faces a strategic inflection point. A regulatory approach that treats models by default as data stores may offer formal security but also constrains innovation and weakens Europe's ability to remain competitive internationally – particularly compared to jurisdictions that assess technical realities more pragmatically. A more precise alignment of oversight with actual model risks, rather than hypothetical data reconstruction scenarios, could therefore become a key competitive factor. If such a shift in perspective succeeds, Europe could develop a regulatory framework that does not dilute protections but strengthens them where genuine harm may occur – while simultaneously placing technological competitiveness more firmly at the center of its AI strategy.

Research Report

Response of Debtors to Rate Changes

HOW BORROWERS RESPOND TO FUTURE CHANGES IN THE INTEREST RATE ON THEIR DEBT MATTERS FOR MONETARY-POLICY TRANSMISSION AND FOR FINANCIAL STABILITY. COMBINING DATA FROM A LARGE BANK, A LETTER RANDOMIZED CONTROLLED TRIAL, AND AN ONLINE SURVEY, WE STUDY THIS QUESTION IN THE CONTEXT OF THE GERMAN MORTGAGE MARKET, WHERE SINCE 2022 BORROWERS HAVE FACED HIGH INTEREST RATES WHEN THEIR RATE FIXATION PERIOD ENDS. WE FIND THAT VARIOUS BORROWER ACTIONS SUBSTANTIALLY REDUCE THE IMPACT OF HIGHER RATES ON MONTHLY PAYMENTS. SURVEY RESPONSES INDICATE HIGH AWARENESS OF THE EVOLUTION OF INTEREST RATES AND CORROBORATE A STRONG PROPENSITY TO PREPARE FOR THE RATE RESET, WHICH WE SHOW EXPERIMENTALLY IS SENSITIVE TO THE SIZE OF THE RATE INCREASE AND TO THE DISTANCE FROM RESET. THE LETTER INTERVENTION DOES NOT AFFECT RATE BELIEFS, CONSISTENT WITH HIGH EX-ANTE INFORMEDNESS AND SELECTION INTO READING, BUT INCREASES AWARENESS OF AVAILABLE OPTIONS AND REFINANCING PROPENSITIES AMONG BORROWERS WHO HAD NOT TAKEN ACTION UNTIL CLOSE TO THE RESET DATE.

Andreas Fuster
Andreas Hackethal
Michael Weber
Virginia Gianinazzi
Philip Schnorpfeil

Introduction

This paper (Fuster et al., 2025) studies how mortgage borrowers prepare for and respond to anticipated changes in the interest rate on their debt. With fixed-rate mortgages, increases in policy rates affect borrowers' monthly payments only when loans reset. Whether households anticipate these higher payments and act in advance determines how quickly monetary policy transmits and how large the risk is of a later wave of financial distress.

We exploit the sharp rise in mortgage rates in Germany after 2021. German mortgages typically amortize over 25–35 years but feature interest-rate fixation periods of 5–15 years. When the fixation period expires, borrowers must refinance at prevailing rates or repay the loan. This institutional structure generates predictable rate resets for large groups of borrowers. Our main message is that German mortgagors are generally well-informed and actively manage rate risk. They adjust contracts, refinance strategically, and build buffers in ways that substantially dampen the pass-through of higher rates into debt-service payments. As a result, we argue that debt-service-driven financial stability risks from the recent rate hikes are limited, and these actions shape monetary policy transmission.

Methodology and Data

The analysis combines three main components linked at the borrower level. First, we use proprietary data from a large German bank. The loan-level data cover about 240,000 mortgagors and nearly 400,000 loans originated between the late 1990s and mid-2025. The data provide annual snapshots of key contract characteristics, as well as some borrower characteristics and records of advisor meetings. The sample focuses on loans with initial fixation of at most 10 years and a substantial remaining balance at reset, with fixation ending between 2018 and 2027.

Second, we run an online survey of the bank's mortgagors. Around 46,000 customers were invited in November 2024 and roughly 2,000 responded. The survey elicits beliefs about mortgage rates, asks borrowers about their preparation for higher future payments, and includes a vignette in which respondents consider a hypothetical refinancing decision under different rate and timing scenarios.

Third, we exploit a large-scale information intervention. In November 2024, the bank sent an information letter to more than 35,000 mortgagors with interest-rate risk, randomly assigning them to different letter variants or to a no-letter control group. The letter explains the rise in mortgage rates, illustrates the effect of a rate increase on monthly payments, and describes options such as partial prepayments and forward mortgages. We track outcomes both in the survey and in the bank data. This design allows us to document real-world behavior, measure awareness and preparation directly, and causally test whether information changes behavior.

Finding 1: Borrowers Actively Manage Refinancing to Smooth Payment Shocks

The administrative data show that borrowers actively manage the timing and terms of refinancing to moderate the impact of higher rates on their monthly payments. If borrowers simply rolled their loans over at the end of the fixation period without any adjustment, the sharp increase in market rates since 2022 would have raised the average monthly payment by roughly 80 euros for loans refinancing after the start of the hiking cycle. In reality, the observed average increase is only about one quarter of that magnitude.

A central tool is the forward mortgage, which

allows borrowers to lock in a future refinancing rate several years before reset, at the cost of a small interest surcharge per year of advance booking. When market rates started to rise quickly in early 2022, the volume of forward mortgages surged, especially among borrowers whose fixation was due to end in one to two years. Once the new higher rate level had become the norm, forward demand declined again. This pattern suggests that borrowers react strongly to the dynamics of rates, not just to their level.

Hazard models of refinancing decisions show a nuanced response. When the prevailing market rate is above the borrower's current rate, the likelihood of refinancing within the same bank falls, reflecting an aversion to locking in a worse rate. At the same time, recent increases in the market rate raise the refinancing hazard, consistent with attention shocks: once borrowers see rates moving up, they become more inclined to act.

Conditional on refinancing, borrowers also adjust contract features to buffer payment changes. In periods of low rates, they often shorten the remaining maturity, keeping monthly payments stable while accelerating principal repayment. After the rate hikes, they become less likely to shorten maturity, thereby limiting the increase in monthly payments. Borrowers also inject additional equity at refinancing somewhat more often when rates are higher than at origination, reducing principal balances. Changes in fixation length provide an additional margin, with a tendency to choose somewhat shorter new fixations after rates have risen. A decomposition exercise attributes most of the dampening of payment increases to the timing of refinancing and to adjustments in maturity, with extra principal repayments playing a smaller but still meaningful role.

Finding 2: Awareness and Preparation for Future Rate Resets Are High

The survey evidence indicates that borrowers are generally well aware of the rise in mortgage rates and that most of them either have already prepared for higher future payments or plan to do so. A large majority report that mortgage rates are substantially higher than a few years ago, and typical perceptions of current rates are close to actual market levels.

Among borrowers with sizeable remaining debts and upcoming rate risk, only a minority report that they are not preparing for the reset. Many in this group in fact face small remaining balances or plan to repay the loan at reset, so that higher rates are less relevant for them. In open text, borrowers mention locking in rates through forward mortgages or related products, making partial or full prepayments, increasing savings to build a buffer, and gathering information by speaking with advisors or comparing offers across lenders.

The vignette experiment complements these findings. In a standardized hypothetical scenario in which a mortgage rate is scheduled to jump noticeably at reset, most respondents state that they would meet an advisor, shop around across banks, and adjust their payments via prepayments or contract changes. When we vary the time until reset, respondents are less likely to plan immediate action when the reset is still three years away rather than one year, especially regarding advisor meetings and income or expenditure adjustments. Anticipation is therefore strong but actions are front-loaded as the reset date approaches. When we vary the size of the rate increase, larger hikes lead to more intended coping via income and spending adjustments, whereas smaller hikes encourage the use of financial options such as forward contracts and longer fixations. Overall, the evidence depicts borrowers as informed, forward-looking, and willing to use several margins to manage debt-service risk.

Finding 3: Letters Move Some Late Movers but Face Selection Constraints

The randomized letter campaign sheds light on the role of formal information interventions. Around one third of treated borrowers recall reading the letter. The letter does not meaningfully change average beliefs about mortgage rates, which were already accurate ex-ante. It does, however, increase familiarity with specific tools such as partial prepayments and forward mortgages, and in the survey it raises the stated likelihood of using some of these instruments.

In the administrative outcomes, the average effect of the letter across all treated borrowers is modest. However, for the subgroup whose fixation ends within six months of the intervention and who have not yet taken action, receiving the letter significantly increases the probability of meeting a bank advisor and of refinancing the loan, including ahead of automatically offered prolongations. The letter thus leads some late movers to act.

At the same time, we document strong selection into reading the letter. Those who read it tend to have higher debts and incomes, higher self-assessed financial literacy, a history of prepayments, and already more accurate beliefs about the rate environment. The intervention therefore mainly reaches the borrowers who were already relatively attentive and sophisticated. This limits the scope for information campaigns to influence beliefs among the least informed borrowers.

Discussion and Conclusion

German mortgagors are generally well informed and actively smooth debt-service shocks, so that realized payment increases are limited. As a result, debt-service-driven financial stability risks from the post-2021 rate hikes appear limited, but the timing and strength of monetary policy transmission depend crucially on these institutional features and borrowers' ability to anticipate and manage future payment changes.

Reference

Fuster, A.; Gianinazzi, V.; Hackethal, A.; Schnorpfeil, P.; Weber, M.:
The Response of Debtors to Rate Changes.
Working Paper, 2025.

Research Report

Using LLMs for Data Engineering in Enterprises Is Challenging

LARGE LANGUAGE MODELS (LLMS) PROMISE TO AUTOMATE DATA ENGINEERING ON TABULAR DATA, OFFERING ENTERPRISES A VALUABLE OPPORTUNITY TO REDUCE COSTLY MANUAL WORK. BUT ADOPTING LLMS IN INDUSTRY BRINGS UNIQUE CHALLENGES, SUCH AS LARGE TABLE SIZES AND THE NEED FOR INTERNAL KNOWLEDGE. OUR ANALYSIS SHOWS THAT LLMS FACE SUBSTANTIAL LIMITATIONS IN REAL-WORLD SCENARIOS, AND OFFERS IDEAS TO SUPPORT BETTER ADOPTION OF LLMS IN INDUSTRY.

Carsten Binnig
Ulf Brackmann
Liane Vogel
Jan-Micha Bodensohn
Anupam Sanghi

Introduction

Large enterprises generate vast amounts of tabular data that drives applications like machine learning and analytical query processing. Data engineering is crucial for understanding this raw data and preparing it for its downstream usage. It encompasses a range of tasks, from data exploration and integration to data transformation and cleaning. Since these tasks often impose significant manual overhead to apply existing tools to the specific data at hand, the automation of individual data engineering tasks like entity matching (Konda et al., 2016) with the help of machine learning has long drawn attention from researchers. Nevertheless, adapting such machine learning approaches to new data sets and tasks often requires computer science expertise, rendering them inaccessible to many practitioners. Recent work has shown that LLMs can be directly applied to data engineering tasks on tabular data, indicating that they achieve state-of-the-art results on various table-based tasks without requiring task-specific architectures and training (Narayan et al., 2022). The out-of-the-box nature of LLMs provides a significant advantage over other machine learning approaches that require supervised training for each data set and task. Therefore, they provide a promising avenue to automate data engineering tasks.

However, data engineering tasks are often more complex than their typical formulations in the scientific community and existing evaluations are built on tables from web sources, which do not fully represent the real-world complexity of tabular data. Therefore in this work, we collaborate with SAP to evaluate the performance of LLMs on data engineering based on data from real enterprise systems.

Enterprise Challenges for LLMs

We aim to take a holistic view of data engineering in enterprises by systematically analyzing the out-of-the-box performance of LLMs, covering enterprise-specific challenges along the dimensions of data, tasks, and knowledge.

The data challenge. Enterprise data differs from public evaluation data sets in various aspects. First, the tables in enterprise databases are substantially larger in their number of columns and rows, and the schema and data are often more complex with column names and values that lack intuitive meaning. Moreover, enterprise data exhibits a much higher sparsity than public data, leaving many cells empty. Together, these factors make data engineering on enterprise data much more challenging.

The task challenge. Beyond the complexities of enterprise data, the enterprise tasks themselves are also more complex. First, while academia typically studies data engineering tasks in isolation, enterprise tasks in practice are often compounds of multiple simpler steps. As such, we argue that the evaluation procedures themselves must change and we need to study accuracy end-to-end to analyze effects such as how errors propagate.

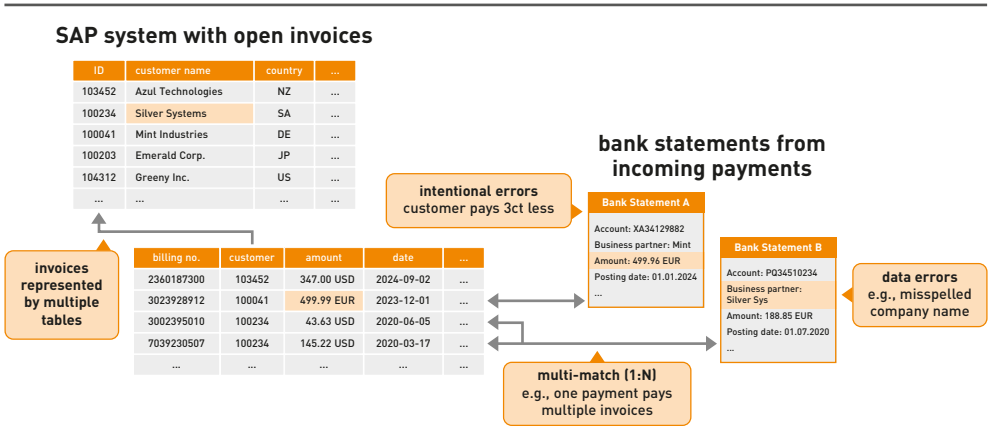

Figure 1: Matching Incoming Payments to Open Invoices

The knowledge challenge. As a last challenge, data engineering in enterprises often requires enterprise-specific knowledge. This is particularly challenging because LLMs will likely not have seen the required information during training, as it is covered only in internal documentation.

Case Study: Enterprise Entity Matching – Incoming Payments & Open Invoices

Entity matching is a data engineering task that often occurs in enterprise scenarios. We study the scenario shown in Figure 1, in which the bank statements of incoming payments must be matched to a company's open invoices.

Study setup. For our case study, we use a data set of payments and invoices following the characteristics of the actual data from an SAP enterprise system and reflecting the challenges described above. The data set contains 15,521 invoices and 12,332 payments. Following existing literature (Narayan et al., 2022), we formulate the entity matching task as a binary classification where the LLM decides if two entities (one payment and one invoice) match; i.e., if the payment pays for the invoice. Note that some of these combinations are part of multi-match cases where one payment pays for multiple invoices or vice versa. To keep the costs tractable, we use a subset of the data set consisting of 790 payment-invoice combinations where the payment does pay for the invoice and 1,210 combinations where it does not. For each combination, we prompt the LLM to predict if it is a match or a non-match. Furthermore, we experiment with presenting the invoices in two formats to the LLM: one format where each invoice is represented by multiple separate tables, and another format where each invoice is represented as one row in a flat view that combines all relevant columns from these tables, mimicking the traditional entity matching setting. We use a few-shot approach that includes one positive and one negative example in the prompt, as zero-shot performance was low on enterprise data. Starting from this simple setup, we make the task incrementally more complex.

Study results. In the following, we explain the different scenarios of increasing task complexity with results provided in Table 1:

	Clean	+ Errors	+ Multi-Matches	+ Multiple Tables
GPT-4o-Mini	0.98	0.58	0.53	0.45
GPT-4o	0.97	0.80	0.64	0.58
Claude 3.5 Sonnet	0.97	0.89	**0.86**	0.58
Llama 3.1 Instruct	**0.99**	**0.95**	0.81	**0.72**

Table 1: Results of Matching Payments to Invoices. Increased Task Complexity Causes a Steady Decrease in F1 Scores

1. (Clean) First, we evaluate on clean 1:1 matches using the flat table views described above, which is closest to the academic setting. In this setting, the payment's memo line includes the correct reference numbers, the payment amount is exactly as stated in the invoice, and the customer name is also identical. As shown in Table 1 (Clean), all models reach very high F1 scores in this setup. Next, we incrementally increase the difficulty based on the challenges we observe in the enterprise scenario.

2. (+ Errors) Data errors in enterprises can differ from those in academic benchmarks, making the matching harder. To demonstrate this, we add representative errors and discrepancies that typically occur in real-world customer data, such as missing or additional digits in the reference numbers or minor discrepancies in the paid amount. As shown in Table 1 (+ Errors), these seemingly small inconsistencies already cause a noticeable drop in accuracy, highlighting the sensitivity of the models to such errors. As shown in Figure 2, discrepancies in textual attributes such as the business partner name result in the most severe performance drops.

3. (+ Multi-Matches) Next, we make the matching even harder by focusing on the multi-match cases, where one payment pays for multiple invoices or multiple payments together pay for one invoice, a setting that is already closer to the real-world scenarios found in enterprises. As shown in Table 1 (+ Multi-Matches), this vastly increases the task's difficulty, since amounts are split between multiple payments and payments might include multiple reference numbers for different invoices.

We further investigate the precision and recall and find that while all models achieve a very high precision close to 1.00, the differences in F1 scores are primarily driven by variations in recall. This indicates that all models take a rather cautious approach, predicting matches only when they are highly certain. As a result, they prioritize high precision in matching and therefore miss many correct matches. While this high precision is sometimes preferred over high recall but low precision, the high rate of

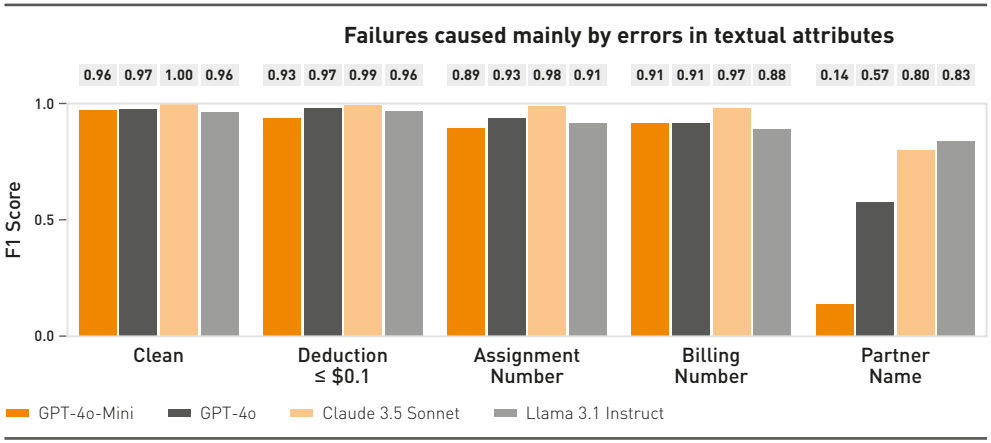

Figure 2: Impact of Different Error Categories When Matching Payments to Invoices

missed matches causes substantial manual efforts for enterprises to find the missing matches.

4. (+ Multiple Tables) Lastly, to show the impact of business entities spanning across multiple tables, we represent invoices using multiple separate tables (as in the original SAP schema) instead of a single flat table view. While metadata about each invoice is stored in one table, specific information like the amount and due date is stored in a second table, and information about customers is stored in another separate customer table. In this scenario, which now closely resembles the actual enterprise task in its full complexity, all models see large performance decreases compared to the previous experiments, indicating that the models have difficulties working with the complex data structures used in enterprises.

Discussion. Through the case study, we demonstrated that automating enterprise tasks with LLMs remains extremely hard. We believe that LLMs out-of-the-box will not be sufficient to address the complexity of enterprise tasks because much of it is inherent to business needs and cannot be avoided, such as having to deal with multi-match cases in entity matching. While reducing complexity by working with database views did improve results, constructing such views causes additional manual overheads since business objects in enterprises can be composed of tens or even hundreds of tables. Instead, complex tasks call for carefully-designed systems that use LLMs as reasoning or tool-calling agents while effectively incorporating human-LLM interactions.

Costs at Enterprise Scale

The cost of using LLMs at enterprise scale is an important orthogonal dimension to the challenges discussed before. In this section, we want to briefly highlight the main cost drivers: the large scale of the data and the high tokenizer fertility (Rust et al., 2021), which is a general problem on tabular data.

Data scale. As described before, enterprise databases are substantially larger than typical data engineering benchmark data sets, with SAP databases often reaching multiple terabytes in size. Table 2 shows how this large size translates into high LLM costs. Processing one gigabyte of tabular data from our SAP corpus with GPT-4o-Mini (OpenAI's cheapest model) already costs about USD 71. Note that this cost does not include generating any output tokens, which are substantially more expensive than input tokens. Nevertheless, the input token costs already provide a rough estimate of the expected high costs. For example, processing the entire customer database with GPT-4o-Mini would cost USD 463 thousand, and using the o1 reasoning model would cost USD 46 million. These high costs clearly prohibit using LLMs on large fractions of the enterprise data.

	Tokens per Byte	USD per GB		
		GPT-4o-Mini	GPT-4o	o1
Wikipedia	0.23	34	574	3,442
SAP CTA (CSV)	0.47	71	1,181	7,085
USD for entire database		462,923	7,715,380	46,292,282

Pricing by OpenAI in February 2025.

Table 2: Tokens per Byte and Cost per GB for Textual and Enterprise Data

Tokenizer fertility. A final important cost driver is the high fertility of existing tokenizers on tabular data. The tokenizer fertility measures the number of tokens required to represent a single piece of text (Rust et al., 2021). Since tokenizers are trained to represent natural language text, their vocabulary contains tokens even for long words. By contrast, tabular enterprise data contains lots of symbols and numerical values. As shown in Table 2, this results in twice the number of tokens required to represent each byte of enterprise data compared to text from Wikipedia, leading to twice the cost per gigabyte. Even if the cost of new models continues to decrease, the costs for processing tabular data will thus remain twice as high. Therefore, future work should focus on creating new tokenizers specifically designed for enterprise data, as well as on creating more efficient textual representations for such data.

Conclusion

In this work, we systematically analyzed the challenges involved in applying LLMs to real-world data engineering scenarios in enterprises. Through experiments with multiple LLMs, we have shown that the effectiveness of these models is substantially affected by enterprise-specific complexities along the dimensions of data, tasks, and knowledge. While the absolute numbers of the observed effects may change with more experiments and newer models, the large gap between enterprise and public scenarios shows the clear need for future research. On the same note, we see a strong need for enterprise-specific LLMs, with many recent activities emerging from research and industry. We hope that our insights and learnings provide a helpful guide and inspiration for future efforts to make LLMs viable for enterprise data engineering and support their adoption in industry.

References

Konda, P.; Das, S.; Suganthan G. C., P.; Doan, A.; Ardalan, A.; et al.:
Magellan: Toward Building Entity Matching Management Systems Over Data Science Stacks.
In: Proceedings of the VLDB Endowment, 9 (2016) 13, pp. 1581–1584.

Narayan, A.; Chami, I.; Orr, L.; Ré, C.:
Can Foundation Models Wrangle Your Data?
In: Proceedings of the VLDB Endowment, 16 (2022) 4, pp. 738–746.

Rust, P.; Pfeiffer, J.; Vulić, I.; Ruder, S.; Gurevych, I.:
How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models.
In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, (2021), pp. 3118–3135.

Insideview

Delivering the EU Equities Consolidated Tape

INTERVIEW WITH EGLANTINE DESAUTEL

Eglantine Desautel
CEO
EuroCTP

Obtaining a comprehensive view of trading activity across EU is currently complex and costly. In response, regulators introduced a unique regulated Consolidated Tape (CT) for each asset class as part of the MiFIR review. EuroCTP was chosen as Europe's inaugural CT Provider for equities and ETFs on December 19th, 2025, following a public tender process managed by ESMA.

How will the CT function in practice?
The EuroCTP system captures pre- and post-trade data from all trading and trade reporting venues across the EU. It aggregates the data and publishes a real-time feed that will provide a consolidated view of all transactions executed in the EU, along with the best bid and offer available at a given time.

What drives the EU's push for a CT in its capital markets?
European markets face the challenge of their own strengths. Their diversity creates complexity. Today, to get a comprehensive view of trading activity across the EU, investors must connect to 200 different markets, each with its own licensing and organized in about 50 technical feeds. This complexity not only makes Europe harder to navigate but it can also give the impression of lower liquidity. Many investors connect to only a subset of trading venues, seeing just a fraction of available liquidity, limiting their positions.

What the CT aims to do is bring the best of both worlds. Preserving the richness of EU markets, while introducing an aggregation layer that effectively creates a virtualized single market from a market data perspective.

Greater participation improves price formation and valuations, creating better conditions for issuers to raise capital. This strengthens fundraising within the EU, helping to finance innovation and drive economic growth.

How will European investors benefit from the CT?
The shares and ETFs CT is a pivotal initiative for Europe, designed to support both institutional and retail investors. For the first time, retail investors will have free, real-time access to consolidated EU market data. Navigating EuroCTP website will also allow them to discover new EU listed instruments, developing cross-border trading.

As the CT receives data at no cost and operates within an efficient structure, industry professionals will be able to obtain aggregated data at highly competitive rates.

What is the timeline for this initiative?
Now selected by ESMA, EuroCTP will initiate the authorization process which aims at obtaining the formal licence.

As the chosen applicant, EuroCTP has started connecting all trading and trade publication venues to ensure readiness for the launch. Work with data redistributors is also underway to secure that they offer the CT feed, real time, as of day one.

We are targeting a go live in July 2026, subject to authorization. Most features are already in place and users can begin testing.

Beyond the regulatory obligation, what makes EuroCTP unique?
As a European player, we are fully committed to delivering a strong CT for the European Union. Our solution was built with the industry to optimize the value for price ratio.

In the United States, the National Best Bid and Offer is determined by capturing a snapshot of available information at the CT level at a specific moment. Conversely, in the European Union, the industry sought a reference value that remains independent of where the CT provider's data centers are located. To meet this requirement, we developed specialized aggregation logic designed for this purpose.

We will provide not only the mandated file formats but also a low-latency feed for users connecting directly to our data center, as well as a cloud-based API for those seeking an easy-to-use interface.

For retail clients, we will offer a tool that allows them to check their trades against the European Best Bid and Offer and other transactions.

The regulation mandates the CT provider to redistribute part of its profit to the data contributors. Because smaller venues can choose not to contribute, we have carefully designed the size of the redistribution envelope to create strong incentives and maximize participation, while keeping our fees highly attractive.

Infopool

News

Annual Conference of the efl in 2025
The efl Annual Conference on "AI 2035: Visions for Business and Society" successfully took place on November 27th, 2025, at Finanz Informatik in Frankfurt. The focus of the conference was on the impact of AI on business and society over the next decade, fostering interdisciplinary dialogue and offering insights from research and real-world applications across sectors and disciplines. We thank Prof. Dr. Oliver Hinz and his team for organizing this year's conference. Our special thanks go to Finanz Informatik for hosting us. For more information, please visit: *https://www.eflab.de/events/efl-annual-conference-2025*.

SFA Best PhD Paper Award for Tino Cestonaro and Niklas Trimpe
efl members Tino Cestonaro and Niklas Trimpe (team Prof. Gomber) won the 2025 SFA (Southern Finance Association) Best PhD Paper Award for their paper "Efficient or not? Price measures in market microstructure". Congratulations!

JAIS Best Paper Award for Prof. Hinz and Co-Authors
Prof. Hinz and co-authors received the JAIS (Journal of the Association of Information Systems) Best Paper Award for their paper "Feedback Loops in Machine Learning: A Study on the Interplay of Continuous Updating and Human Discrimination". Congratulations!

BAI Science Award for Dr. Bender
Dr. Micha Bender (team Prof. Gomber) won the BAI (Bundesverband Alternative Investments) science award for his dissertation "The Impact of Regulation on Securities Trading". The aim of his thesis is to deepen our understanding of the effects of regulation on market quality and price discovery. Thereby, he contributes to the current academic debate and offers valuable insights into the complex interactions between market rules and market behavior. Congratulations!

IHK Dissertation Award for Dr. Jürgensmeier
Dr. Lukas Jürgensmeier (team Prof. Skiera) won the IHK (Industrie- und Handelskammer) award for his dissertation "Measuring Fair Competition". The aim of his dissertation is to examine how fair competition – particularly on digital platforms such as Amazon – can be measured. As part of this research, for example, he develops a methodology to assess whether recommendations on digital platforms lead to unfair competition. Congratulations!

Prof. Binnig Associate Editor at VLDB 2025
Prof. Binnig has been recognized as a Distinguished Associate Editor for VLDB 2025, the 51st International Conference on Very Large Data Bases in London, September 1st to September 5th, 2025. This recognition highlights his contributions to the data management community.

Prof. Skiera Area Editor at Journal of Marketing
Prof. Skiera has been appointed Area Editor at the Journal of Marketing, the leading academic journal in the field of marketing and the one with the highest impact score. This prestigious role reflects his outstanding contributions to marketing research and his recognition within the global scholarly community.

Selected efl Publications

Bucher-Koenen, T.; Hackethal, A.; Koenen, J.; Laudenbach, C.:
Gender Differences in Financial Advice.
Forthcoming in: American Economic Review.

Dündar, U.; Abou Nabout, N.; Skiera, B.:
Getting the Pace Right: Performance of Budget Allocation Heuristics in Online Advertising.
Forthcoming in: International Journal of Research in Marketing.

Hackethal, A.; Hanspal, T.; Hartzmark, S. M.; Bräuer, K.:
Educating Investors About Dividends.
Forthcoming in: The Review of Financial Studies.

Jakobs, J.; Clapham, B.; Muntermann, J.:
Empirical Evidence on Misconduct and Law Enforcement in Crypto Asset Markets.
In: 8th Cryptocurrency Research Conference (CRC2025); Athens, Greece, 2025.

Jürgensmeier, L.; Skiera, B.:
Measuring Self-Preferencing on Digital Platforms.
Forthcoming in: Journal of Marketing.

Leuz, C.; Meyer, S.; Muhn, M.; Soltes, E. S.; Hackethal, A.:
Who Falls Prey to the Wolf of Wall Street? Investor Participation in Market Manipulation.
Forthcoming in: Management Science.

Mihale-Wilson, C.; Felka, P.; Hinz, O.:
Mobile ICT Outages and Public Safety: Is there a Digital Divide in Terms of Safety?
Forthcoming in: European Journal of Information Systems.

Miller, K.; Lukic, K.; Skiera, B.:
Impact of the General Data Protection Regulation (GDPR) on Online Tracking.
Forthcoming in: International Journal of Research in Marketing.

Ozcan, F.; Chung, Y.; Chronis, Y.; Gan, Y.; Wang, Y.; et al.:
LLMs and Databases: A Synergistic Approach to Data Utilization.
In: IEEE Data Engineering Bulletin, 49 (2025) 1, pp. 32–44.

Schnorpfeil, P.; Weber, M.; Hackethal, A.:
Inflation and Trading.
In: Electronic Markets, 173 (2025), Article No. 104166.

von Zahn, M.; Liebich, L.; Jussupow, E.; Hinz, O.; Bauer, K.:
Knowing (Not) to Know: XAI and Human Metacognition.
Forthcoming in: Information Systems Research.

Yang, C. L.; Bauer, K.; Li, X.; Hinz, O.:
My Advisor, Her AI and Me: Evidence From a Field Experiment on Human-AI Collaboration and Investment Decisions.
Forthcoming in: Management Science.

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

Infopool

RESEARCH PAPER:
DOES RANSOMWARE MAKE INVESTORS "WANNACRY"? ON INVESTORS' DIVERGENT REACTIONS TO RANSOMWARE HITS AND NEAR MISSES

In recent years, ransomware has become one of the most severe cyber threats, causing major disruptions and financial losses. Rationally, investors should respond negatively to ransomware disclosures, yet this is not always observed. Drawing on norm theory, this paper examines a paradox in which investors react negatively to ransomware hits but positively to near misses – events where disruptions were narrowly avoided. This outcome bias leads investors to perceive near misses as positive, even though they still reflect security failures. Using an event study on ransomware disclosures by US public firms, this paper estimates abnormal stock returns and finds evidence consistent with this bias. Ransomware hits led to stock price drops of 4.40%, while near misses were rewarded with gains of 2.87%. These findings highlight how investors may misinterpret near-miss events as signs of resilience, overlooking the underlying cybersecurity risks they reveal.

Schuetz, S. W.; Chen, Y.; Forderer, J.; Ma, Y.
In: MIS Quarterly, 49 (2025) 3, pp. 1153–1168.

RESEARCH PAPER:
GENERATIVE AI AT WORK

The authors study the staggered introduction of a generative AI-based conversational assistant using data from 5,172 customer-support agents. Accessing AI assistance, the results show that less experienced and lower-skilled workers improve both the speed and quality of their output, while the most experienced and highest-skilled workers see small gains in speed and small declines in quality. The authors also find evidence that AI assistance facilitates worker learning and improves English fluency, particularly among international agents. While AI systems improve with more training data, they show that the gains from AI adoption are largest for moderately rare problems, where human agents have less baseline experience but the system still has adequate training data.

Brynjolfsson, E.; Li, D.; Raymond, L.
In: The Quarterly Journal of Economics, 140 (2025) 2, pp. 889–942.

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, 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