"Show Me What's Wrong!":
Combining Charts and Text to Guide Data Analysis

Beatriz Feliciano*   Rita Costa†   Jean Alves‡   Javier Liébana§   Diogo Duarte¶   Pedro Bizarro‖
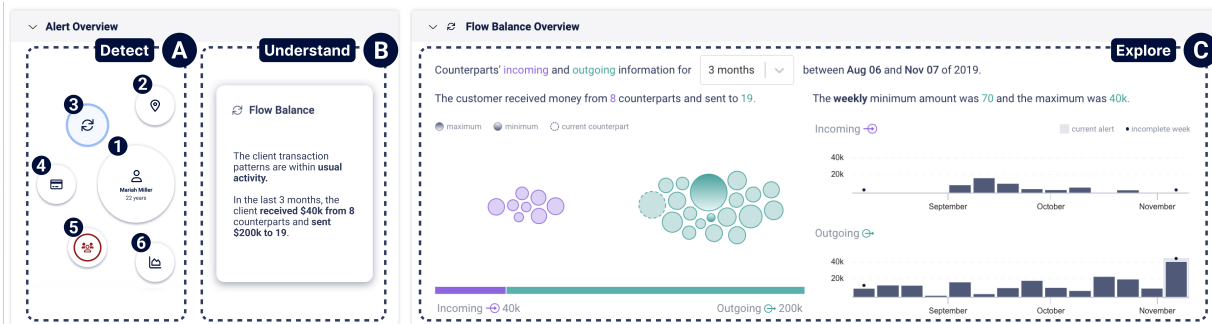
Feedzai

Figure 1: Proposed interface with multiple levels of detail exploration composed of (A) a region where the alert is segmented in the subgroups that compose it (A.1, A.2, A.3, A.4, A.5, and A.6) and where groups that require more attention (in this case, A.5) are highlighted in red; (B) an automatically generated text summary of a selected area (A.3) that provides a broad understanding of the group; and (C) an interactive graphical representation of all the data points of the selected area to explore information in detail.

ABSTRACT

Analyzing and finding anomalies in multi-dimensional datasets is a cumbersome but vital task across different domains. In the context of financial fraud detection, analysts must quickly identify suspicious activity among transactional data. This is an iterative process made of complex exploratory tasks such as recognizing patterns, grouping, and comparing. To mitigate the information overload inherent to these steps, we present a tool combining automated information highlights, Large Language Model generated textual insights, and visual analytics, facilitating exploration at different levels of detail. We perform a segmentation of the data per analysis area and visually represent each one, making use of automated visual cues to signal which require more attention. Upon user selection of an area, our system provides textual and graphical summaries. The text, acting as a link between the high-level and detailed views of the chosen segment, allows for a quick understanding of relevant details. A thorough exploration of the data comprising the selection can be done through graphical representations. The feedback gathered in a study performed with seven domain experts suggests our tool effectively supports and guides exploratory analysis, easing the identification of suspicious information.

Index Terms: Human-centered computing, Visualization, Data exploration, Visualization design and evaluation methods.

1 INTRODUCTION

Data exploration for risk analysis is a crucial task in multiple domains such as financial loan approval [33], medical diagnosis [1],

---

*e-mail: beatriz.feliciano@feedzai.com
†e-mail: rita.costa@feedzai.com
‡e-mail: jean.alves@feedzai.com
§e-mail: javier.liebana@feedzai.com
¶e-mail: diogo.duarte@feedzai.com
‖e-mail: pedro.bizarro@feedzai.com

and financial fraud detection [18]. Analysts must, among other tasks, identify data patterns and compare a current event (e.g., loan application, health exam parameters, or performed transaction) with historical information (e.g., previous loans, patient history, or transactions through time). For this, domain experts need to carry out a non-linear and iterative data exploration process [6] based on tabular data, which makes exploring information and detecting events a complex, fatiguing, and time-consuming task [25].

Although a common problem across domains, our work focuses on the detection of financial fraud – an issue that resulted in the loss of 159 billion dollars in the USA in 2023 [30]. To better understand how the current analysis process is carried out in this use case, we conducted observation sessions with eleven financial fraud analysts. This enabled us to define design goals tailored to experts' specific needs, which informed our proposal outline.

We observed that, when reviewing a financial fraud alert, analysts come across hundreds or even thousands of data points that they need to explore and analyze to reach a conclusion. Adding to the complexity of the issue, this process must be carried out rapidly [17], in a few minutes or even seconds [34], through the exploration of tabular data [25]. Analysts frequently come across multiple data aggregations and need to switch from the overall view to a more granular exploration of a given high-level concept (e.g., geographical information). These changes in granularity frequently lead to a loss in overall context. Additionally, financial analysts often work alongside automated systems designed to aid them in their decision making [25] by providing a risk estimate in the form of a risk-score, and even by performing risk-attribution, identifying which variables contribute to an elevated risk-score, if any. This adds further complexity to the task, as analysts are not only required to explore the data themselves but also to incorporate the automated system's insights into their judgment.

To aid in these issues and leverage human perception, current research proposes different solutions, from data summarization – through graphical representations [2, 8, 12, 15, 18, 22, 24, 27, 29, 32, 34, 40] or textual descriptions [14, 28, 36, 37] – to information highlighting [4, 7, 35] and multiple levels of detail exploration [20, 24, 25, 26, 34, 39]. These techniques alone impose some limitations

on the data evaluation, such as information overload [10].

We present an innovative system, including a novel user interface promoting time-efficient and highly-informed decision-making. This is achieved by combining a multi-modal summary of the available data in textual and visual representations. Furthermore, the tool integrates the insights from fraud detection methods, leveraging them to produce representations that convey the key risk-assessment insights provided by automated systems. As presented in Figure 1, the interface is composed of three different elements: (A) a console for guiding analysts by presenting each area of analysis and allowing them to **detect** which contribute to an elevated risk-assessment; (B) a **Large Language Model (LLM)** generated text summary containing the relevant information and automated system insights for the selected analysis area, to help them **understand** its characteristics and why it is particularly risky or not; (C) an interactive graphical representation of all the data points of the selected area that varies and is adjusted according to its attributes, letting users further **explore** the data. Each interface element (from left to right) has an increased level of detail and can be seen by users at any time, preventing context loss and aiding information retrieval. The text acts as the link between the high-level representation of the data and the more complex and detailed charts.

To evaluate our system we performed a qualitative assessment for a financial fraud detection use case with seven domain experts. The interface was praised for its ease of use and integration into existing workflows, reducing the current analysis workload and allowing a faster insights retrieval. The console (A) and text summary (B) received particular acclaim for their effectiveness in providing concise summaries of alerts and the areas that compose them, thereby eliminating the necessity to examine all data points in simpler cases.

The key contributions of our work are:

- Identification of the characteristics of exploratory analysis in the context of a financial fraud use case based on observation sessions conducted with eleven fraud analysts;

- An interactive interface tailored for exploring and identifying anomalies in large multi-dimensional financial datasets, guiding the analysis and pointing out relevant areas through automated textual and visual insights comprising different levels of detail, which are displayed side-by-side;

- A user study with seven financial fraud detection specialists validating the usefulness of the designed interface.

2 RELATED WORK

This section provides a brief review of prior research on data exploration and high-risk event identification with a focus on the ease of information retrieval by domain experts.

2.1 Graphical representation of attributes

Event detection is an analytics problem that requires visual exploration [11]. Consequently, it is common to graphically represent this multi-level data and its attributes [17, 18].

Previous research presents various approaches that facilitate behavior and event detection by visually encoding each relevant data attribute through different graphical representations adjusted to the attribute characteristics and the insights that users are looking for [18, 40, 27, 32]. In the field of fraud detection, systems such as EVA [18] and FraudAuditor [40] introduce dashboards that represent multiple relevant data attributes. Despite taking advantage of the human perception system [17], these approaches present all features simultaneously, which can be overwhelming for analysts.

Another frequently researched technique to explore multi-dimensional datasets and gather activity patterns is the use of glyph representations [2, 12, 22, 24], which encode various data attributes in a unique visualization. In financial use cases, approaches such as VaBank [24] do so by individually representing each transaction as an element and encoding each of its features by adding multiple visual representations to it. While these approaches reduce the amount of graphs that users need to evaluate, they increase the visualization complexity, as they require the memorization of multiple elaborated visual encodings. Specifically in the field of fraud detection, where analysts are not accustomed to evaluating data in graphical format, it is essential to design simple visualizations [18], as verified in the conducted user study (refer to Section 6).

Additionally, research also proposes systems that perform automatic recommendations of graphical representations according to the data attributes [8], subsets of data attributes [15] or even to inputted user intentions [29]. However, these techniques are inadequate for time-constrained use cases, as they require manually editing the generated charts for adequate insight retrieval.

To address the aforementioned issues, our system balances the amount of displayed information and the visualization complexity, providing relevant insights without overwhelming users.

2.2 Text summary

The use of textual descriptions of data and charts is a growing technique used to communicate insights such as outliers, data values, extremes, and trends [14]. By reading a text description summarizing the data at hand, users are able to immediately retrieve key takeaways, thus having an early advantage in the analysis [36].

There are several possible approaches to generating these textual insights. One of those is the usage of LLMs to automatically generate this type of content. There are benefits to this technique, namely in producing content that conveys meaningful information about data and its distribution. Nevertheless, despite its usefulness, these insights are more commonly generated by LLMs, which can output inaccurate information (*hallucinations* [28, 37]). Consequently, the unchecked use of LLMs introduces the possibility that the text summary contains misleading or deceptive conclusions, which, in turn, can harm the analyst's information retrieval.

To minimize the issues described above, state-of-the-art approaches emphasize the importance of providing users with the necessary means to confirm the validity of the information present in the description [14, 36] – for example, by allowing access to tabular or graphical representations of the dataset. Our proposed tool includes a Hallucination Detection system (described in Section 5.2), to further mitigate the possibility of generating erroneous insights.

2.3 Multiple levels of detail

To avoid data overload and facilitate information digestion, certain techniques propose dividing information into layers and presenting details gradually to users. This allows for an evaluation of summarized [20, 25, 34, 24] or grouped [26, 39] data, and for exploration, in full detail, of smaller portions of the information.

Current research focuses mainly on the use of this technique combined with visual analytics approaches, but we believe that users would also benefit from high-level insights provided in text format, as these have proved to support initial data exploration [36]. Our system supports textual insights, which are described in further detail in Sections 4.2 and 5.2, combining them with highlighting and exploration techniques to further empower users.

2.4 Highlight relevant information

Highlighting important information, be it within text [4, 35], tables [18], or through annotations within charts [7], is a way of directing users' attention to specific details. In the context of data exploration, information highlighting can guide the analysis process. For example, in EVA [18] transactional data concerning fraudulent activity is highlighted in red. However, if the wrong information is brought to attention, there is a potential to mislead analysts or delay

their decision-making. As such, when combining highlighting techniques with insights that may contain incorrect information – like the ones generated by LLMs –, interfaces should be accompanied by an explanation of the highlighting identification process [14]. Our proposal tackles this issue by providing a textual explanation – detailed in Sections 4.2 and 5.2 – which gives users an understanding of the reasons for highlighting while working as a summary that helps to quickly gather the main analysis takeaways.

3 DESIGN GOALS

To derive design objectives for this research, we recruited eleven financial fraud analysts from our network. We conducted one-hour remote observation sessions with each of them, to perceive their data review process. The analysts' work consists of reviewing transactions alerted as possibly fraudulent by an automated system and, in a matter of seconds, assessing whether they correspond to illicit activity. To this end, they iteratively check information regarding the currently flagged transaction and compare it with the alerted entity's past activity, evaluating if they are in line with each other and whether the alert corresponds to an anomalous event.

We were able to determine that, to perform this assessment, analysts divide data exploration into different subsets, henceforth referred to as *Knowledge Areas* (KAs), – for example, the alerted person's demographics, past alerted transactions, transactional activity, or the counterpart's demographics – and explore each area at a time in a non-linear way. By dividing exploration into KAs, experts are able to reduce the laboriousness of the analysis and more easily identify unusual activity.

Currently, the data exploration process is entirely based on tabular data. This requires analysts to invest a significant amount of time navigating through tables to gather insights on relevant patterns and identify suspicious areas, adding complexity to the task.

Based on the insights resulting from the conducted sessions and the literature review, we identified the following design goals:

- **(DG1) Identification of KAs requiring more attention**. Users should be able to easily assess all the different areas that compose the data and to quickly understand which ones might require more careful analysis.

- **(DG2) Summarization of key information within a KA**. Users should easily retrieve the main takeaways of each KA: the mean of the values, trends (ascending or descending), extremes, and differences, among others.

- **(DG3) Detailed exploration of the attributes of a KA**. All the data related to each KA should be present in the system, allowing users to gather detailed concept-specific information;

- **(DG4) Comparison between an individual data point** (currently alerted transaction) **and groups of data points** (historic of transactions). Users should be able to compare the currently examined event with all the data points in the dataset.

4 VISUALIZATION DESIGN

To structure the interface, we propose a segmentation of the data into KAs. The following KA overviews are considered for financial fraud detection: (A.1) alerted person, (A.2) location, (A.3) flow balance, (A.4) card, (A.5) counterpart, and (A.6) alerted person activity. Even though we have six KAs in our proposal, the tool's design can accommodate up to 12 areas (but, to avoid information overload, we recommend care in the considered number of KAs).

As illustrated in Figure 1, our proposal is composed of three elements: (A) a Knowledge Area Console, which is the starting point of the analysis, where the alert is segmented and the analyst is guided to the most relevant KAs, (B) an LLM generated text summary for the selected KA, which states the most pertinent information of the chosen area and where its relevance for the analysis can

be assessed, and (C) a graphical representation of all the data of each KA, enabling a detailed analysis for specific cases. Since they provide high-level insights of the alert data, the Knowledge Area Console (A) and the text summary (B) elements are grouped in an area called Alert Overview. The graphical representation element (C) is part of an area with the name of the selected KA, as it consists of detailed information about the chosen subset.

Given analysts' proficiency in exploring tabular data [18], our tool includes a data table below the presented approach. This is not reflected in the interface images as it corresponds to the current analysis format and does not consist of a novel proposal.

4.1 Knowledge Area Console (A)

The Knowledge Area Console is the starting point of the analysis and where experts can quickly identify the areas of the data that might require more careful analysis (DG1). It consists of a holistic visual depiction of the data where each concept is represented with an icon. There are as many icons as KAs, which are manually designed by experts depending on the use case. Here, users can select the KA for which the tool provides the automated text summary (B) and the graphical representation (C).

Since the alerted person is the focus of the review for the financial fraud analysts, we chose to represent it as the central element of the Console. This KA corresponds to demographic data and insights about past alerts of the alerted person. Other KAs are placed around it, as they all derive from the alerted person's activity.
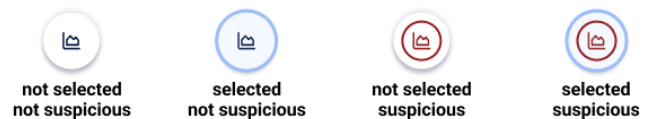
Figure 2: Different states that a KA can have in the Console region.

As shown in Figure 2, KAs have different representations depending on their selection and risk status. We highlight, with a red circle, KAs that are identified as containing potentially suspicious behavior, thus requiring a more careful examination. We describe implementations on how to identify which KAs are suspicious in Section 5.1. This highlight is the first visual cue that communicates to the analyst that something might be out of the ordinary in the data, such that risk analysts interviewed by us claim that, with this design, they would know what is wrong almost immediately.
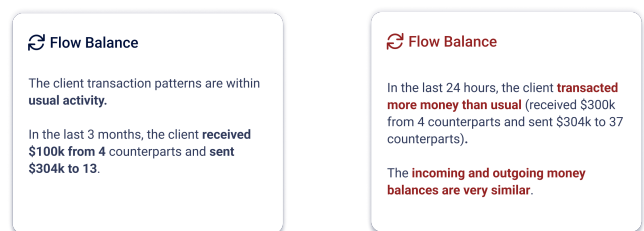
4.2 Knowledge Area Text Summary (B)

Figure 3: Examples of text summaries for different KAs: alerted person (on the left) and alerted person activity (on the right).

This area serves as the link between the high-level data representation (A), described in Section 4.1, and the detailed insights (C), discussed in Section 4.3. Analysts resort to this part of the interface to quickly gather the most relevant insights about the selected KA (DG2) and, in case of identified suspicious information, understand what was marked as risky activity by the system. The main goal

of the text summary is to succinctly outline the key values pertaining to the KA, incorporating the insights provided by the automated system. The user study described in Section 6 highlighted that, if the information gathered in this area answers all analysts' questions, no further details are examined and analysis time is saved.

As illustrated in Figure 3, the focus of each KA summary is adjusted not only to the data about the subset but also to the insights that the analysts identified during observation sessions (see Section 3) as being most important for the subset in question. Besides stating the main attributes of the area (for the alerted person KA, for example, the person's name, age, and country are represented), it also describes how the current event compares to the alerted entity's behavioral patterns (DG4). Based on those patterns, the system's risk-assessment methods (described in Sections 5.1 and 5.2) evaluate if the event is an outlier.
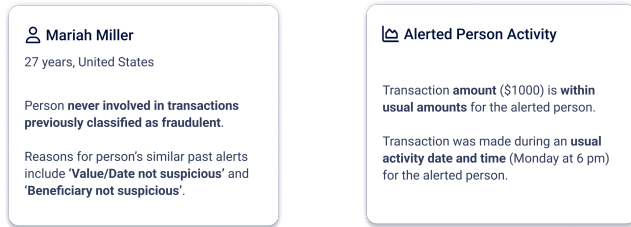

Figure 4: Example of a text summary of a KA not flagged as fraudulent (on the left) and of a KA flagged as fraudulent (on the right).

Should the summary contain any piece of information indicating that the currently alerted transaction constitutes suspicious activity, that area of the text is highlighted in red, as represented in Figure 4. Just like the summary, its text highlights – bold and red color – are automatically generated using an LLM, described in Section 5.2.

4.3 Knowledge Area Graphical Representation (C)

In this portion of the interface, all the information that composes a KA is graphically represented (DG3). This area is designed for use cases where an exhaustive evaluation of the data is required. It allows analysts to confirm the automated insights provided in the text summary (B) area, as it has been shown that analysts can be critical of erroneous automated insights [3]. Here, the analyst can visualize sets of graphs designed to represent the historical data of the selected KA. Consider, for instance, the flow balance KA (refer to Figure 1), where it is important for analysts to gather insights about the total incoming and outgoing amounts, the transactional activity evolution through time, and the counterparts involved in it. Here, their goal is to understand if the balance between incoming and outgoing amounts or involved counterparts is suspicious [13].

Figure 5 lists some examples of different KAs' graphical representations. Following previous research [18], we gave preference to familiar visualizations – such as bar charts, histograms, stacked bars, and area plots – to reduce the analysts' learning curve and thus the information retrieval time. Even though we designed a visual depiction for each data subset to leverage the retrieval of insights, we will not cover in further detail their gathered insights and visual encodings, as these representations are not the focus of this paper.

Each graphical representation is accompanied by a short subtitle that provides a brief summary of the properties of the graph – count of elements, mean, max and min values, and so forth. Information about the currently alerted transaction is highlighted in the graphical representations with a gray background, easing its comparison with the depicted history of data (DG4). Similarly, if the subset that constitutes a KA contains data pertaining to a previously confirmed fraudulent event, that information is highlighted in red, as shown in the alerted person overview of Figure 5.
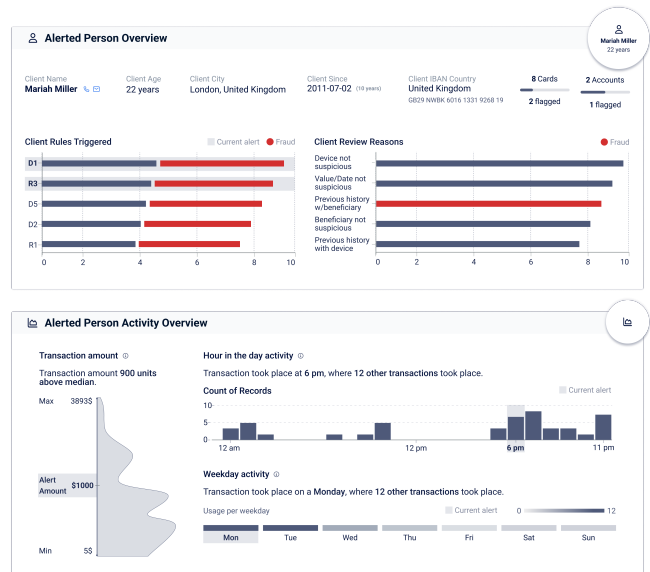

Figure 5: Examples of graphical representations of different KAs: alerted person (on top) and alerted person activity (on the bottom).

5 BACKEND SYSTEM

In this section, we describe how the backend system is designed, from the input data to the visual and textual summary generation. In Figure 6 we present a schematic representation of the tool's architecture.

5.1 Automated Risk-Assessment and Attribution

First, the current event data is fed into the system, being stored in the Database component, which contains data from previous transactions. Then, it is fed into the Knowledge Area Parser, which maps the information into KAs. Due to the specific challenges and properties of different use cases (e.g. transaction fraud, bank account opening fraud), domain experts must pre-define each KA and the parsing methods to segment the data into these fields.

After segmentation, the input data is fed into the Intelligence Component, which produces two outputs for each KA: a risk-score, quantifying the overall risk associated with an event; and, if the score is elevated according to a predefined threshold, a risk-attribution – an identification of the top-k variables responsible for the elevated risk-score. This information is then used in the Knowledge Area Console (A) representation, described in Section 4.1.

For the current implementation, we utilized an existing rule system comprised of a set of user-defined rules that encode expert knowledge on what constitutes risky behavior. Should a rule be triggered, the KAs associated with the variables that caused said rule to be triggered are flagged as a source of risk. This service may also check a user or device blocklist, in which case, the justification for blocking said entity (i.e., relevant KAs) is saved to the Database at the time of addition to said blocklist. Nevertheless, our approach could be used in combination with other risk assessment methods, such as **Machine Learning (ML)** models, and in that case we could use well-established attribution explanation methods such as Shapley values [21] or LIME [31] to identify the key variables responsible for an elevated risk-score.

5.2 Summary Generation

The processed data, along with the outputs of the Intelligence Component, are then fed to the Summary Generator, which is responsible for producing each KA text summary (B), described in Sec-
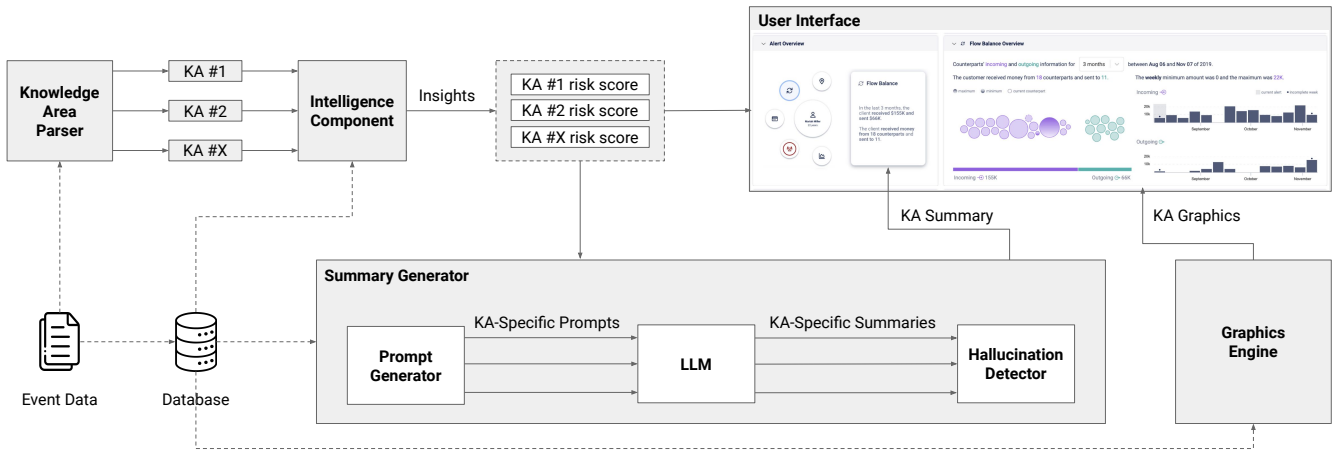
Figure 6: Overview of the architecture of the system.

tion 4.2. This sub-system follows a Retrieval-Augmented Generation (RAG) [19] approach, comprising three components: (1) The Prompt Generator, responsible for giving proper context, as well as introducing and framing the data and the Intelligence Component insights adequately for use by (2) an LLM, which generates the textual summary based on the provided data and context, and (3) the Hallucination Detector system, which ensures that no factually incorrect data is included in the produced summary.

This structure ensures that, via the generated prompt, all the necessary data and context is fed to the LLM, along with the Intelligence Component's insights (e.g., whether or not a KA is suspicious). This approach aims to ensure that the LLM is only tasked with converting the provided information into a human-readable format, in the form of a short paragraph. Similar RAG approaches have been shown to lead to improvement of the generated content while reducing the likelihood of hallucinations [5].

Guaranteeing that the provided summaries are trustworthy is of extreme importance, as a faulty summary could potentially harm the analysts' decision process. For further mitigation of hallucination risk, we include the Hallucination Detector which employs faulty generation detection methods (e.g., SelfCheckGPT [23]) and, in the case of an identified hallucination, guarantees that sub-par generations are not shown by refusing to provide a summary. For these rare cases, the information *"Summary not available"* is displayed instead. As an alternative solution, we could explore the usage of a dynamic adjustment of the values of generation parameters, for example, by decreasing the value of top-p – guaranteeing that only the highest probability tokens are sampled – for the token decoding process. This has been shown to promote factuality in the model's output [16], while reducing its variability and increasing the likelihood of repetition in the text generation.

5.3 Graphics Engine

The Graphics Engine receives as input all the data concerning the activity of the alerted person to generate a set of format-appropriate visual representations for each KA, described in Section 4.3.

Due to the lack of methods to automatically create visual representations without the need for manual adjustments (see Section 2.1), this approach demands a higher level of domain-expert input, requiring graphs to be designed and implemented. After that, each KA graphical representation is fed with its relevant attributes.

6 USER STUDY

To validate the interface, we conducted a user study with seven (P1-P7) of the eleven financial fraud detection analysts who took part in

the observation sessions (see Section 3).

Data. We used a dataset with transactional synthetic data, similar to the one proposed by Jesus et al. [9], which was designed to capture the dynamics of fraud in bank account transactions over time, providing a realistic representation of the data encountered in real-world scenarios. The features included in the dataset cover customer information, and transaction details, among others.

Procedure. We conducted one-hour semi-structured interviews with each participant via Zoom. They were briefed about the purpose of the system, how they could interact with it, and its visual encodings. After that, they were given remote access to a computer with the tool (Figure 1) and were asked to freely interact with it and explore the data for all the KAs while thinking aloud [38]. They were informed during the sessions that they could ask design or functionality-related questions to the researchers.

Findings. The feedback from the user study proves the usefulness of our tool in aiding data exploration and insight retrieval in multi-dimensional datasets, in the context of financial fraud detection. While P2 and P6 shared some concerns about the automatically highlighted and summarized information potentially introducing bias to the analysis process, all users expressed interest in adopting this tool and emphasized several key benefits of using it.

Experts noted how **this approach would improve their efficiency**. P1 mentioned that, by looking into the Knowledge Area Console (A) and the subsets' text summary (B), they could, *"in a matter of two seconds, understand what the data is showing"* and added that they were able to classify the alert as fraudulent or not *"almost immediately"*. P3 stated that the multiple levels of detail exploration almost eliminates the need to see data in a tabular format, which would be necessary only in very specific cases where all details need to be considered. Experts believe that, for the vast majority of the cases, the text summary and graphical representation of the data provide all the required granular insights.

Furthermore, P4 highlighted that this tool "would enable the most basic review to be almost automatic, allowing [analysts] to dedicate [their] time and effort to more advanced cases".

Additionally, participants **praised the tool's capabilities in conveying behavioral patterns**. P6 stated that it enabled them to *"get an overview of the behaviors of the alerted person"* for each analysis area, and P4 emphasized that, with this system, they could more efficiently gather *"context about the alerted entity's activity and transaction"* than they currently do through tabular data.

Users also mentioned that **the interface effectively pointed to alert subsets that had potentially risky information**. The visual representations – particularly the use of colors in the Knowledge

Area Console (A) to indicate potentially risky activity and the KA text summary (B) to explain the risk classification – were highly praised. Users noted that their analysis was driven to highlighted areas, which helped them navigate and focus on the relevant data subsets more efficiently. P3 commented that *"it is easier to identify what is wrong or what is not common to the person's activity"*. P6 added that the tool enabled them to *"easily identify risky areas"* and to quickly understand which unusual information they comprised.

Lastly, participants **appreciated how all the information was accessible in the same location**. Analysts valued having all the necessary information in one place, with the different levels of detail always visible. P2 mentioned that since *"everything is in the same place, there is no need to scroll and fewer clicks are needed"*, as they do not have to constantly order and filter data. P1 added that this would minimize actions not directly related to retrieving information, leaving more time to analyze data and gather insights.

While the goal of this study was to validate the interface as a whole and not the specific graphical representations of each KA (C), we find it important to highlight that analysts had difficulties interpreting and tended to avoid more complex or less known graphs – such as the bubble chart shown in Figure 1. This indicates that the detailed visual depiction of some KAs will require adjustments.

7 CONCLUSION AND DISCUSSION

Detecting anomalous activity in transactional datasets requires navigating and interpreting complex and multi-dimensional data fast and accurately. Despite that, users often suffer from information overload, which leads to performance issues in information retrieval. The proposed tool addresses this issue by dividing data into subsets, highlighting the ones that might require more urgent analysis, and providing a multi-modal summary of the available data per area, both in LLM-generated text and in one or multiple charts. This aids users in simultaneously gathering high-level and granular insights from the data, enabling them to identify patterns, group, and compare information. The performed user study has shown that our solution promotes time-efficient and highly informed decision-making, contributing to reduced work and lower cognitive overload in the fraud analysis process. For future work, we aim to adjust the graphical representation of some analysis subsets and to perform formal user tests to quantitatively measure the contributions of our proposal. Furthermore, we plan on testing the usage of ML models for the risk assessment of the data and further study on hallucination mitigation methods for producing the text summary.

REFERENCES

[1] R. H. Alvi, M. H. Rahman, A. A. S. Khan, and R. M. Rahman. Deep learning approach on tabular data to predict early-onset neonatal sepsis. *Journal of Information and Telecommunication*, 5(2):226–246, 2021. 1

[2] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):280–289, 2016. doi: 10.1109/TVCG.2015.2467196 1, 2

[3] M. De-Arteaga, R. Fogliato, and A. Chouldechova. A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020. 4

[4] T. Doi. Effective highlighting modes of graphical user interfaces in visual information search. *Universal Access in the Information Society*, 22(1):111–119, jul 2023. 1, 2

[5] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023. 5

[6] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Commun. ACM*, 55(4):45–54, apr 2012. doi: 10.1145/2133806. 2133821 1

[7] A. K. Hopkins, M. Correll, and A. Satyanarayan. Visualint: Sketchy in situ annotations of chart construction errors. *Computer Graphics Forum*, 39(3):219–228, jul 2020. doi: 10.1111/cgf.13975 1, 2

[8] K. Hu, D. Orghian, and C. Hidalgo. Dive: A mixed-initiative system supporting integrated data exploration workflows. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA '18. Association for Computing Machinery, New York, NY, USA, jun 2018. doi: 10.1145/3209900.3209910 1, 2

[9] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. P. Ribeiro, J. Gama, and P. Bizarro. Turning the Tables: Biased, Imbalanced, Dynamic Tabular Datasets for ML Evaluation. *Advances in Neural Information Processing Systems*, 2022. 5

[10] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. *Visual Analytics: Scope and Challenges*, pp. 76–90. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-71080-6_6 2

[11] J. Kielman, J. Thomas, and R. May. Foundations and frontiers in visual analytics. *Information Visualization*, 8(4):239–246, dec 2009. doi: 10.1057/ivs.2009.25 2

[12] C. Kinkeldey, J.-D. Fekete, and P. Isenberg. BitConduite: Visualizing and Analyzing Activity on the Bitcoin Network. In A. P. Puig and T. Isenberg, eds., *EuroVis 2017 - Posters*. The Eurographics Association, 2017. doi: 10.2312/eurp.20171160 1, 2

[13] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri. Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review. *IEEE Access*, 9:82300–82317, 2021. doi: 10.1109/ACCESS.2021.3086230 4

[14] P.-M. Law, A. Endert, and J. Stasko. Characterizing automated data insights. In *2020 IEEE Visualization Conference (VIS)*, pp. 171–175, oct 2020. doi: 10.1109/VIS47514.2020.00041 1, 2, 3

[15] D. J.-L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. Parameswaran. Avoiding drill-down fallacies with vispilot: assisted exploration of data subsets. IUI '19, p. 186–196. Association for Computing Machinery, New York, NY, USA, mar 2019. doi: 10.1145/3301275.3302307 1, 2

[16] N. Lee, W. Ping, P. Xu, M. Patwary, P. N. Fung, M. Shoeybi, and B. Catanzaro. Factuality enhanced language models for open-ended text generation. *Advances in Neural Information Processing Systems*, 35:34586–34599, 2022. 5

[17] R. A. Leite, T. Gschwandtner, S. Miksch, E. Gstrein, and J. Kuntner. Visual analytics for event detection: Focusing on fraud. *Visual Informatics*, 2(4):198–212, dec 2018. doi: 10.1016/j.visinf.2018.11.001 1, 2

[18] R. A. Leite, T. Gschwandtner, S. Miksch, S. Kriglstein, M. Pohl, E. Gstrein, and J. Kuntner. Eva: Visual analytics to identify fraudulent events. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):330–339, 2018. doi: 10.1109/TVCG.2017.2744758 1, 2, 3, 4

[19] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020. 5

[20] D. Liu, S. Alnegheimish, A. Zytek, and K. Veeramachaneni. Mtv: Visual analytics for detecting, investigating, and annotating anomalies in multivariate time series. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1), apr 2022. doi: 10.1145/3512950 1, 2

[21] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017. 4

[22] C. MaÇãs., E. Polisciuc., and P. Machado. Glyphsome: Using som with data glyphs for customer profiling. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - IVAPP*, pp. 301–308. INSTICC, SciTePress, 2020. doi: 10.5220/0009178803010308 1, 2

[23] P. Manakul, A. Liusie, and M. J. Gales. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. *arXiv preprint arXiv:2303.08896*, 2023. 5

[24] C. Maçãs, E. Polisciuc, and P. Machado. Vabank: Visual analytics for banking transactions. In *2020 24th International Conference Information Visualisation (IV)*, pp. 336–343, 2020. doi: 10.1109/IV51561.2020.00062 1, 2

[25] C. Maçãs, E. Polisciuc, and P. Machado. Atovis – a visualisation tool for the detection of financial fraud. *Information Visualization*, 21(4):371–392, 2022. doi: 10.1177/14738716221098074 1, 2

[26] X. Mu, K. Xu, Q. Chen, F. Du, Y. Wang, and H. Qu. MOOCad: Visual Analysis of Anomalous Learning Activities in Massive Open Online Courses. In J. Johansson, F. Sadlo, and G. E. Marai, eds., *EuroVis 2019 - Short Papers*. The Eurographics Association, 2019. doi: 10.2312/evs.20191176 1, 2

[27] P. H. Nguyen, R. Henkin, S. Chen, N. Andrienko, G. Andrienko, O. Thonnard, and C. Turkay. Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):77–86, 2020. doi: 10.1109/TVCG.2019.2934609 1, 2

[28] J. Obeid and E. Hoque. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model. In *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 138–147. Association for Computational Linguistics, oct 2020. doi: 10.18653/v1/2020.inlg-1.20 1, 2

[29] A. Pandey, A. Srinivasan, and V. Setlur. Medley: Intent-based recommendations to support dashboard composition. *IEEE Transactions on Visualization and Computer Graphics*, 29(01):1135–1145, jan 2023. doi: 10.1109/TVCG.2022.3209421 1, 2

[30] R. Ramaiya and S. Rogers. A colossal $159 billion lost to fraud and scams, finds new research from Feedzai and the Global Anti-Scam Alliance, February 2024. 1

[31] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016. 4

[32] J. Rzeszotarski and A. Kittur. Crowdscape: interactively visualizing user behavior and output. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, p. 55–62. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2380116.2380125 1, 2

[33] M. A. Sheikh, A. K. Goel, and T. Kumar. An approach for prediction of loan approval using machine learning algorithm. In *2020 international conference on electronics and sustainable communication systems (ICESC)*, pp. 490–494. IEEE, 2020. 1

[34] P. Silva, C. Maçãs, E. Polisciuc, and P. Machado. Visualisation tool to support fraud detection. In *2021 25th International Conference Information Visualisation (IV)*, pp. 77–87, 2021. doi: 10.1109/IV53921.2021.00022 1, 2

[35] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498, jan 2016. doi: 10.1109/TVCG.2015.2467759 1, 2

[36] N. Sultanum and A. Srinivasan. Datatales: Investigating the use of large language models for authoring data-driven articles. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 231–235, oct 2023. doi: 10.1109/VIS54172.2023.00055 1, 2

[37] B. J. Tang, A. Boggust, and A. Satyanarayan. Vistext: A benchmark for semantically rich chart captioning. *arXiv preprint arXiv:2307.05356*, jun 2023. 1, 2

[38] M. Van Someren, Y. F. Barnard, and J. Sandberg. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress*, 11, 1994. 5

[39] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins. #fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, 2014. doi: 10.1109/TVCG.2014.2346922 1, 2

[40] J. Zhou, X. Wang, J. Wang, H. Ye, H. Wang, Z. Zhou, D. Han, H. Ying, J. Wu, and W. Chen. Fraudauditor: A visual analytics approach for collusive fraud in health insurance. *IEEE Transactions on Visualiza-tion and Computer Graphics*, 29(6):2849–2861, 2023. doi: 10.1109/TVCG.2023.3261910 1, 2