Received: 24 February 2023 | Accepted: 15 June 2025

DOI: 10.1111/1911-3846.13070

CAAA ACPC CONTEMPORARY ACCOUNTING RESEARCH RECHERCHE COMPTABLE CONTEMPORAINE

ARTICLE

The informational content of key audit matters: Evidence from using artificial intelligence in textual analysis

Stephan Küster[1] | Tobias Steindl[2] | Max Göttsche[3]

[1]School of Business and Economics, Department of Finance, Accounting and Taxation, Freie Universität Berlin, Berlin, Germany

[2]University of Regensburg, Regensburg, Germany

[3]Catholic University of Eichstätt-Ingolstadt, Ingolstadt, Germany

Correspondence
Stephan Küster, School of Business and Economics, Department of Finance, Accounting and Taxation, Freie Universität Berlin, Thielallee 73, D-14195 Berlin, Germany.
Email: stephan.kuester@fu-berlin.de

Abstract

This study provides empirical evidence that key audit matters (KAMs) are informative for future negative accounting outcomes. We employ FinBERT—a deep learning model designed for natural language processing that allows human-like text comprehension—to demonstrate that goodwill-related KAMs are predictive of firms' future impairments. Our findings reveal that utilizing KAMs as a stand-alone predictor for future impairments provides meaningful predictive power. By exploring the semantic content of reported KAMs, we find that their predictive power is primarily driven by text passages covering how both the firm and the auditor exercise judgment in the accounting and auditing of goodwill. Furthermore, we show that KAMs are incrementally predictive beyond several firm-level determinants and disclosures in annual reports. Finally, our additional analyses indicate that (1) KAM-predicted impairment probabilities are relevant to capital markets, (2) KAMs are useful for predicting the magnitude of goodwill impairments, and (3) the predictive power extends to other KAM topics. Collectively, our findings enhance the understanding of the informational content of KAMs, which is a key rationale for their introduction.

KEYWORDS

audit reporting, FinBERT, goodwill impairment, key audit matters, natural language processing, prediction

Accepted by Matthew Lyle.

CAAA ACPC CONTEMPORARY ACCOUNTING RESEARCH RECHERCHE COMPTABLE CONTEMPORAINE

Le contenu informatif des questions clés de l'audit : données issues de l'utilisation de l'intelligence artificielle dans l'analyse textuelle

Résumé

Cette étude fournit des données empiriques démontrant que les questions clés de l'audit (QCA) sont informatives pour les résultats comptables négatifs futurs. Nous utilisons FinBERT, un modèle d'apprentissage profond conçu pour le traitement du langage naturel qui permet une compréhension du texte similaire à celle des humains, afin de démontrer que les QCA liées à l'écart d'acquisition sont prédictives des pertes de valeur futures des entreprises. Nos résultats révèlent que l'utilisation des QCA comme prédicteur indépendant des pertes de valeur futures offre un pouvoir prédictif significatif. En explorant le contenu sémantique des QCA divulgués, nous constatons que leur pouvoir prédictif est principalement déterminé par les passages du texte qui décrivent la manière dont l'entreprise et l'auditeur exercent leur jugement dans la comptabilité et l'audit de l'écart d'acquisition. En outre, nous montrons que les QCA ont un pouvoir prédictif supplémentaire qui va au-delà de plusieurs déterminants au niveau de l'entreprise et de la communication d'informations dans les rapports annuels. Enfin, nos analyses supplémentaires indiquent que (1) les probabilités de perte de valeur prédites par les QCA sont pertinentes pour les marchés financiers, (2) les QCA sont utiles pour prédire l'ampleur des pertes de valeur de l'écart d'acquisition, et (3) le pouvoir prédictif s'étend à d'autres thèmes des QCA. Collectivement, nos conclusions améliorent la compréhension du contenu informatif des QCA, qui est l'une des principales raisons de leur introduction.

MOTS-CLÉS

divulgation d'audit, FinBERT, perte de valeur de l'écart d'acquisition, prédiction, questions clés de l'audit, traitement du langage naturel

1 | INTRODUCTION

The audit report serves as the primary communication tool between auditors and the public. However, its informational content has faced criticism for relying on standardized "boilerplate" language and structure that often convey symbolic rather than communicative value (e.g., Church et al., 2008). For instance, during the 2008 financial crisis, critics contended that simple pass-or-fail judgments failed to provide useful information for predicting potential risks, prompting calls for more firm-specific audit information (e.g., Mock et al., 2013).

In response, standard setters worldwide have expanded audit reports for publicly listed firms to include key audit matters (hereafter, KAMs). KAMs outline firm-specific financial statement risks and how auditors address them. For each identified KAM, the auditor details (1) considerations regarding inherent and control risks in the description section, and (2) the auditor's response to the resultant detection risk in the response section. One objective of KAMs is to inform stakeholders about areas involving significant managerial judgments (e.g., ISA 701), as well as the implications of these judgments for future accounting outcomes.

To understand whether KAMs achieve this objective, it is essential to establish whether the disclosed information assists stakeholders in predicting future accounting outcomes. Such predictive ability offers direct, non-market-based evidence of their informational content (Minutti-Meza, 2021). Prior literature has primarily focused on the aggregated capital-market effects of KAM disclosure, yielding mixed results (e.g., Bédard et al., 2019; Gutierrez et al., 2018; Porumb et al., 2021). Consequently, the predictive power of KAMs remains an open question. It also remains unclear which type of information within KAMs holds the highest informational content and whether these disclosures are incrementally predictive (Burke et al., 2023; Seebeck & Kaya, 2023).

This study aims to address these gaps by exploring three research questions: (1) Is the informational content within goodwill-related KAMs predictive of future goodwill impairments? (2) Which type of information within goodwill-related KAMs possesses the highest predictive power? (3) Are goodwill-related KAMs incrementally predictive of future goodwill impairments beyond firm disclosures and firm-level determinants?

We focus on goodwill-related KAMs due to the inherent risks in goodwill valuation (ISA 315.A7). Impairment decisions are influenced by both economic factors and managerial incentives, making goodwill accounting a domain characterized by significant managerial judgment (e.g., Han et al., 2021). Therefore, impairments present an ideal context in which KAMs should enhance understanding. In addition, goodwill is frequently reported as a KAM due to its complexity and risk of material misstatement (e.g., Burke et al., 2023; Financial Reporting Council [FRC], 2016; Klevak et al., 2023). Thus, evidence supporting the predictive power of goodwill-related KAMs strengthens the case for the overall informational content of KAMs.

We compile a novel data set of 6,277 European firm observations from 2016 to 2021, linking goodwill-related KAMs to subsequent goodwill impairments. Using a machine learning approach, we employ FinBERT, a deep learning model for natural language processing (NLP), to predict future impairments based on KAM text (e.g., Huang et al., 2023; Siano & Wysocki, 2021). FinBERT, which has been trained on over 800,000 financial documents, captures contextual meaning in financial language and outperforms traditional “bag-of-words” methods used in prior accounting and auditing research (Huang et al., 2023). This allows for a deeper understanding of KAMs’ informational content.

To test our first research question, we fine-tune the FinBERT algorithm using a portion of our labeled data set, enabling it to identify patterns in KAMs that differentiate between goodwill impairments and non-impairments in the subsequent period. We then apply the resulting model to a holdout subsample to generate predicted impairment probabilities, which we use to evaluate the stand-alone predictive power of goodwill-related KAMs. Technically, we test whether KAMs outperform a random classifier. Our results show that the full KAM text, as well as the description and response sections individually, possess significant predictive power for future goodwill impairments. The full KAM text outperforms a random classifier by 32.7 percentage points, followed by the description section (29.8 percentage points) and the response section (29.1 percentage points). These findings confirm KAMs’ stand-alone predictive power, supporting the notion of their informational content.

To address our second research question, we conduct a “masking test” by altering or removing specific KAM text sequences to assess the predictive power of particular information types within KAMs. Prior literature suggests the importance of both numbers and contextual information in textual analysis (e.g., Huang et al., 2023; Siano & Wysocki, 2021). We find that removing numbers reduces predictive power by up to 6%, while randomizing words results in a 23% drop—indicating that context plays a larger role than raw numbers. We further explore the semantic meaning of KAMs by deleting sentences that capture different information types defined under International Standard of Auditing (ISA) 701. These results suggest that the predictive power of goodwill-related KAMs is primarily driven by information regarding how management and auditors exercise judgment in accounting for and auditing goodwill.

To address our third research question, we examine the incremental predictive power of KAMs beyond firm disclosures. Focusing on the description section—which may reiterate information from the annual report—we use FinBERT word embeddings and cosine similarity to calculate the semantic similarity between KAM sentences and related goodwill disclosures in firms' annual reports. Deleting similar sentences from the KAM does not reduce predictive power, suggesting that the description section of the KAM is not merely a summary of information already disclosed in the annual report. We also train a separate FinBERT model using goodwill-related firm disclosures and find that KAMs significantly outperform it in predicting impairments. Next, we assess the incremental predictive power of KAMs beyond other firm-level determinants identified in the literature as predictors of goodwill impairments. We train a Gradient Boosting machine learning algorithm using firm-level variables to predict future impairments and compare its predictions with those of the KAM prediction model. Again, KAMs yield significantly higher predictive power. When combining all predictors, KAMs add approximately 8% to the predictive power of a benchmark model based on both goodwill-related firm disclosures and determinants. These findings indicate that KAMs possess not only stand-alone predictive power but also incremental predictive power beyond firm disclosures and determinants.

We perform several additional tests to substantiate our main findings. First, we examine whether KAM-predicted impairment probabilities are relevant to the capital market. While our analysis reveals that these probabilities are value relevant across firms of all sizes, we observe a statistically significant association with short-term price reactions only in smaller firms. This finding is in alignment with the FRC's assertion that KAMs are particularly beneficial for investors in smaller firms (FRC, 2016). Second, we explore whether goodwill-related KAMs can predict the *magnitude* of future impairments. KAMs explain approximately 26% of the variation in goodwill impairment magnitudes, primarily driven by the response section, suggesting they are informative not only regarding the occurrence of future impairments but also concerning their magnitude. Third, we investigate generalizability by using a limited data set of goodwill-related critical audit matters (CAMs) from US firms, where extended audit reports became effective only in 2019. The prediction results show stand-alone predictive power, suggesting that our findings extend beyond the European context. Fourth, we explore the predictive power for future accounting outcomes of other KAM topics that require significant judgment. For example, we analyze KAM text sequences related to going concern to predict future going-concern opinions. Results show comparable predictive power to our main model, indicating that our findings extend beyond goodwill-related KAMs.

This study makes three significant contributions to the literature. First, we enhance the growing body of research that empirically investigates the informational content of KAMs (e.g., Bédard et al., 2019; Gutierrez et al., 2018; V. Li & Luo, 2023; Porumb et al., 2021). While these market-based studies focus on aggregated capital-market effects of KAM disclosures, we investigate the informational content of KAMs primarily from a non-market perspective. Unlike market-based studies that provide indirect evidence for the claim that KAMs have informational content (Minutti-Meza, 2021) due to confounding effects of other disclosures (e.g., Myers et al., 2018), we provide direct empirical evidence regarding the informational content of KAMs. Specifically, we are the first to demonstrate that goodwill-related KAMs are informative for future goodwill impairments from a non-market perspective. Moreover, our methodology focuses on what is being disclosed in KAMs rather than how. Therefore, our content-based analysis adds to the literature by providing more fine-grained evidence on the semantic meaning of KAM disclosures and their relation to future goodwill impairments. We also demonstrate the incremental informational content of KAMs beyond established firm-level variables and information already disclosed in annual reports, complementing prior market-based studies that have not consistently found evidence of such incremental informational content (Lennox et al., 2023). By complementing our main non-market perspective with an additional market-based analysis, we further show that the informational content of KAMs is associated with investor reactions in a nuanced manner: while goodwill-related KAMs provide incrementally new information to investors of smaller firms, investors of larger firms appear to have already anticipated this information before its publication. By integrating these perspectives, we offer a more comprehensive understanding of KAMs' informational content.

Second, we contribute to the growing accounting literature utilizing textual analysis (e.g., Bochkay et al., 2020; Davis & Tama-Sweet, 2012; Dyer et al., 2017; Huang et al., 2018). Prior studies mainly rely on "bag-of-words" approaches (Loughran & McDonald, 2016), which ignore grammar, word order, and context. In contrast, our study aligns with a nascent stream of research utilizing advanced textual analysis techniques, employing contextualized word embeddings (e.g., Bingler et al., 2022; S. V. Brown et al., 2024; Siano, 2022; Song, 2021). This deep learning approach more accurately captures meaning in financial texts (Araci, 2019; Huang et al., 2023; Yang et al., 2020). While prior studies have used large language models to predict accounting outcomes (e.g., Siano, 2022), our study is, to the best of our knowledge, the first to leverage FinBERT to assess the predictive power of KAMs in forecasting negative accounting outcomes like impairments. In doing so, we address calls for more advanced textual analysis methods in accounting research (Bochkay et al., 2023).

Third, we enrich the goodwill accounting literature (e.g., Beatty & Weber, 2006; Glaum et al., 2018; Han et al., 2021; Ramanna & Watts, 2012). While prior studies highlight firm fundamentals and managerial incentives as key predictors of goodwill impairment, we show that audit risk—the rationale underlying KAMs—is also incrementally predictive of future goodwill impairments. This illustrates the role of extended audit reports as a supplementary information source for assessing the likelihood of future impairments.

2 | BACKGROUND AND RESEARCH QUESTIONS

2.1 | Relevance and informational content of KAMs

Following the 2008 financial crisis, the expectation gap between the public perceptions of auditor performance and actual auditor responsibilities widened (Ruhnke & Schmidt, 2014). First, public trust eroded as auditors issued unqualified opinions for entities that later went bankrupt (e.g., Sikka, 2009). Second, audit reports, which provided a simple pass-or-fail judgment at that time, were criticized for lacking detail on financial statement risks, critical managerial judgments, and the auditors' work (e.g., European Commission, 2011). Thus, the demand for more client-specific audit information increased (Mock et al., 2013).

In response to this expectation gap, standard setters worldwide have enhanced auditor report content to increase its usefulness (e.g., FRC, 2013; IAASB, 2015; PCAOB, 2017). For instance, Regulation No. 537/2014 of the European Union (EU) mandates that auditors provide an extended report for the financial statement audits of public interest entities. One significant component of this extended report is KAMs, defined as "those matters that, in the auditor's professional judgment, were of most significance in the audit of the financial statements of the current period" (ISA 701.8).[1] KAMs typically highlight areas of heightened financial statement risk, including the risk of material misstatement, significant judgments, accounting estimates, and significant events or transactions (ISA 701.9). If an auditor does not identify a KAM, this must also

[1]In the EU, the reporting of KAMs is defined in the International Standard on Auditing (ISA) 701, "*Communicating Key Audit Matters in the Independent Auditor's report*." This standard became effective for audits of financial statements ending on or after December 15, 2016.

be disclosed (ISA 701.16). It is crucial to emphasize that KAMs are not a substitute for original information provided by the firm itself or for modified auditor opinions. Instead, they offer disclosures on selected topics intended to help users better understand the client-specific financial statement risks, judgments, and the auditors' work (ISA 701.2).

KAMs generally comprise two sections (ISA 701.13). First, auditors must describe the identified matter and why it represents a KAM for the current audit. This description section includes references to the matter's disclosure in the financial statements and considerations of inherent and control risks that render it significant to the audit (ISA 701.A42–45). Second, auditors should also explain how they addressed the identified matter(s) in their financial statement audit. This response section details detection risk and includes aspects of the auditors' approach, audit procedures performed, outcomes of those procedures, key observations, and a conclusion on the matter (ISA 701.A46–51). Collectively, these client-specific KAM disclosures aim to enhance financial statement users' understanding of the auditors' work and the significant risks and judgments present within the financial statements.

Prior research on the informational content of KAMs can be categorized into two main streams. The first stream examines the impact of KAMs on individual behavior in experimental settings. These studies generally conclude that KAMs can provide valuable information for financial statement users. Specifically, KAMs appear to direct investor attention to critical sections in financial statements (Sirois et al., 2018), facilitate information intake by complementing financial statement disclosures (Dennis et al., 2019), influence assessments of auditor responsibility (Brasel et al., 2016; Gimbar et al., 2016; Kachelmeier et al., 2020), and enhance the credibility of reported information and the audit itself (Elliott et al., 2020; Moroney et al., 2021). Regarding investment decisions, Christensen et al. (2014) and Rapley et al. (2021) found that nonprofessional investors were more likely to alter their investment decisions when presented with extended audit reports containing KAMs compared to standard reports lacking such disclosures. Focusing on goodwill-related KAMs, Köhler et al. (2020) discovered that professional investors consider content changes within KAMs for their investment decisions, whereas nonprofessional investors do not. Additionally, Boolaky and Quick (2016) conducted an experiment with German bank directors and determined that credit lending decisions were not significantly influenced by KAM disclosures. This indicates that the informational content of KAMs may vary across different financial statement user groups. Overall, this body of literature supports the assertion that KAMs are informative.

The second stream of literature comprises studies that investigate the aggregate capital-market effects of KAMs. In contrast to the first stream, empirical evidence from these archival studies raises questions about the informational content of KAMs. Some studies focus on the first-time adoption of KAM reporting, revealing that the introduction of KAMs enhanced the disclosure of risk-related information and improved the readability, evaluative content, and visual guidance within audit reports (Seebeck & Kaya, 2023; Smith, 2023). While evidence suggests the value relevance of initial KAM reporting (Klevak et al., 2023), studies typically find no significant association with investor reactions, as measured by abnormal returns and trading volume across various jurisdictions (Bédard et al., 2019; Burke et al., 2023; Gutierrez et al., 2018; Lennox et al., 2023; Liao et al., 2022; Seebeck & Kaya, 2023). This indicates that while auditors may report relevant information within KAMs, it does not provide sufficient new information pertinent to investors (Lennox et al., 2023). However, in the context of debt markets, Porumb et al. (2021) found evidence of reduced loan spreads following the introduction of KAM reporting, suggesting that KAMs are informative for loan contracting purposes. Other studies examine whether variations in KAM disclosure influence aggregated capital-market effects, yielding mixed results. Critics of the new legislation expressed concern that KAMs could lead to excessive "boilerplate" reporting (IAASB, 2013; Minutti-Meza, 2021; Mock et al., 2013); nonetheless, evidence indicates that KAM disclosures vary based on firm-specific and auditor-specific characteristics (Abdelfattah et al., 2021; Bepari et al., 2022; Honkamäki et al., 2022; Klevak et al., 2023; Rousseau & Zehms, 2024) and that these variations may be informative for investors. For example, Klevak et al. (2023) reported negative market reactions associated with a higher number of disclosed KAMs. Burke et al. (2023) show that the introduction of CAMs led to changes in firm disclosure practices, but that there is no consistent evidence of subsequent reactions in capital markets following CAM adoption. Conversely, Liao et al. (2022) find only weak evidence that variations in KAM type, length, and client-specific focus are associated with abnormal returns. Overall, this stream of literature presents mixed evidence regarding the informational content of KAMs, leaving it uncertain whether KAMs achieve one of their intended objectives: to inform stakeholders about financial statement areas associated with significant managerial judgments.

2.2 | Research questions

Our research questions examine the informational content of KAMs in financial statement areas involving significant managerial judgment. We specifically focus on goodwill, which is frequently reported as a prominent KAM in extended audit reports and reflects a significant area of managerial judgment (Burke et al., 2023; FRC, 2016; Klevak et al., 2023; H. Li et al., 2019). Under IFRS, goodwill accounting follows an impairment-only approach, where management regularly needs to test the recoverability of the recorded goodwill. This impairment test involves considerable managerial discretion regarding forecasting future cash flows and determining measurement assumptions, particularly concerning the discount rate that impacts the impairment decision (e.g., Filip et al., 2015). In line with the inherent managerial judgment, there is empirical evidence that goodwill impairment is not only determined by economic factors but also by managerial incentives (e.g., Beatty & Weber, 2006; Han et al., 2021; Ramanna & Watts, 2012), corporate governance (Gunn et al., 2018; Kabir & Rahman, 2016), and public enforcement (e.g., Glaum et al., 2018). If goodwill-related KAMs assist financial statement users in assessing the likelihood of future goodwill impairments, we conjecture that this predictive power indicates that KAMs are informative regarding significant judgments inherent in goodwill accounting.

While prior research is largely silent on the predictive power of KAMs for financial statement areas of significant judgment, two arguments support the notion that goodwill-related KAMs can predict future goodwill impairments. First, auditors detail audit risk components for each specific matter. For goodwill, inherent risk descriptions may inform stakeholders about external and client-specific economic conditions relevant to impairment likelihood. Since managerial incentives can influence impairment decisions, control risk disclosures may further indicate impairment risk. Additionally, auditors likely expand procedures when the margin between recoverable amount and book value of goodwill is narrow, suggesting that detection risk information can enhance user evaluations of future goodwill impairments.

Second, prior literature indicates that KAM disclosures capture client-specific uncertainties and risk factors. For instance, Camacho-Miñano et al. (2024) demonstrate that more comprehensive KAM disclosures, which can signal future economic developments, are associated with higher audit fees. Additionally, studies reveal that KAM disclosures vary based on certain risk-related firm characteristics (Bepari et al., 2022; Klevak et al., 2023). Thus, we anticipate that auditors may convey higher impairment risks in goodwill-related KAMs.

Taken together, these arguments suggest that goodwill-related KAMs are predictive of future goodwill impairments. However, it remains unclear whether auditors have incentives to enhance the transparency of KAM disclosures. This is important because auditors can comply with the ISA 701 requirements by reporting "boilerplate" KAMs. Currently, there is no evidence linking higher transparency in KAM disclosures to greater auditor reputation, remuneration, or market share (Minutti-Meza, 2021). Therefore, the possibility remains that auditors may perceive KAMs as a compliance exercise. Hence, we propose the following research question:

> **Research Question 1 (RQ1).** Are goodwill-related KAMs predictive of future goodwill impairments?

KAM disclosures comprise a description section that addresses inherent and control risk considerations and a response section that provides insights into the level of detection risk. For each section, ISA 701 recommends specific information categories that auditors can disclose. These categories focus on diverse topics and may yield insights into the predictive power of individual information types. Thus, our second research question reads as follows:

> **Research Question 2 (RQ2).** Which type of information within goodwill-related KAMs possesses the highest predictive power?

Several scholars argue that KAMs may not provide new information to financial statement users, but rather reiterate information contained in the client's annual report (Gutierrez et al., 2018; Lennox et al., 2023; Liao et al., 2022).[2] Additionally, prior research has identified various firm characteristics that account for goodwill impairments (e.g., Filip et al., 2015; Han et al., 2021). Therefore, it remains unclear whether KAMs provide incremental information beyond (1) disclosures in the firm's annual report and (2) significant firm-level determinants identified in previous literature that explain future goodwill impairments (Amel-Zadeh et al., 2020). Hence, our third research question reads as follows:

> **Research Question 3 (RQ3).** Are goodwill-related KAMs incrementally predictive of future goodwill impairments beyond firm disclosures and firm-level determinants associated with goodwill impairment?

3 | METHODOLOGY

3.1 | Use of the large language model FinBERT

To investigate our research questions, we employ a transfer learning approach for NLP using the FinBERT language model developed by Huang et al. (2023).[3] For text classification, FinBERT's neural network can be fine-tuned with a labeled data set to identify patterns in text sequences associated with specific categories (e.g., negative or positive sentiment). Once fine-tuned, the model can analyze new, unseen text sequences and assign probabilities to their respective categories. In this study, we fine-tune FinBERT to differentiate between goodwill impairments and non-impairments for a firm in year $t + 1$, based on the KAM text inputs from year t. The resulting model assigns a probability of future goodwill impairment to each observation.[4]

3.2 | Data and sample selection

We construct our labeled data set by merging KAM text data from Audit Analytics Europe with corresponding goodwill data from Refinitiv. We focus on Europe for two reasons: most

[2]A related but distinct line of research also suggests that KAM disclosures influence firm disclosures (Andreicovici et al., 2020; Jahan & Karim, 2024). Although exploring this connection is outside the scope of this study, it emphasizes the importance of examining the semantic similarity between the goodwill-related KAM disclosures and the financial statement disclosures to determine whether KAM disclosures are incrementally predictive for future goodwill impairments.

[3]Detailed information regarding the rationale for using FinBERT and the implementation of the machine learning algorithms can be found in Appendix S1 in the Supporting Information. For an excellent review of transfer learning approaches and large language models in general, see Bochkay et al. (2023) and Siano and Wysocki (2021).

[4]Supplementary Appendix S2 presents an illustrative example of our prediction approach using FinBERT.

TABLE 1 Sample selection.

	Observations		
	Total	***thereof impairment***	***thereof non-impairment***
KAM data (2016–2021) from Audit Analytics Europe	54,315		
Less: KAMs not relating to goodwill	(38,546)		
Goodwill-related KAMs	6,328		
Less: Missing data for goodwill-related KAMs	(51)		
Starting sample	**6,277**	**961**	**5,316**
Less: 20% randomly chosen **(testing sample)**	**(1,255)**	**(211)**	**(1,044)**
Remaining sample	5,022	750	4,272
Less: 5% randomly chosen **(validation sample)**	**(251)**	**(38)**	**(213)**
Less: Randomly exclusion of non-impairments	(3,347)	–	(3,347)
Training sample	**1,424**	**712**	**712**

Note: This table reports the sample selection procedure for the testing, validation, and training samples used for our prediction models.

firms apply IFRS (IASB, 2022), ensuring consistent goodwill measurement, and audit firms report KAMs under ISA 701 for fiscal years ending on or after December 15, 2016, providing a uniform reporting framework for auditors.

Table 1 illustrates our sample selection process. We first obtain all KAM text data from the Audit Analytics Europe database for the years 2016–2021. Starting with 54,315 observations, we exclude all KAM topics unrelated to goodwill (38,546 observations) and remove all KAMs with missing data (51 observations).[5] We then merge the remaining KAMs with the firms' corresponding goodwill impairment data for the next fiscal year from Refinitiv, which forms our starting sample of 6,277 observations.[6] It contains 961 impairment observations (15.3%), which is comparable to other studies (e.g., Glaum et al., 2018).

Our prediction models require testing, training, and validation samples, as shown in Figure 1. Following accepted machine learning practice, we randomly select a testing sample of 1,255 observations (i.e., 20% of the starting sample). This sample will be used to perform our main analyses—that is, to test and evaluate our prediction model. Before testing, we fine-tune the FinBERT model using training and validation samples. The training sample enables the model to learn KAM classification, while the validation sample ensures proper model specification and avoids overfitting. Because our labeled data is imbalanced, with more firms not recognizing goodwill impairments than those that do, we remove observations to ensure a balanced data set for training purposes (i.e., random undersampling). This is important, as an unbalanced data set would likely result in a prediction model that classifies most observations as non-impairments to achieve high prediction accuracy. Instead, we want the model to learn patterns in KAMs indicative of future goodwill impairments.[7] After randomly selecting 251 observations (i.e., 5% of the starting sample, excluding the testing sample) for our validation sample, we randomly exclude 3,347 non-impairment observations, resulting in a balanced training data set of 1,424 observations (i.e., training sample).

[5]We rank-order KAM topics with respect to their frequency and find that goodwill-related KAMs (i.e., topics labeled "goodwill" and "goodwill and intangible assets") rank second, surpassed only by revenue recognition.

[6]When a firm reports a positive net goodwill in its balance sheet, we assume that goodwill impairment is zero when data is missing. We test the robustness of our label in Section 6.

[7]It is important to note that our main results are based on the unbalanced testing sample that is randomly drawn from our starting sample. The balanced training and validation sample is only required for the fine-tuning process.

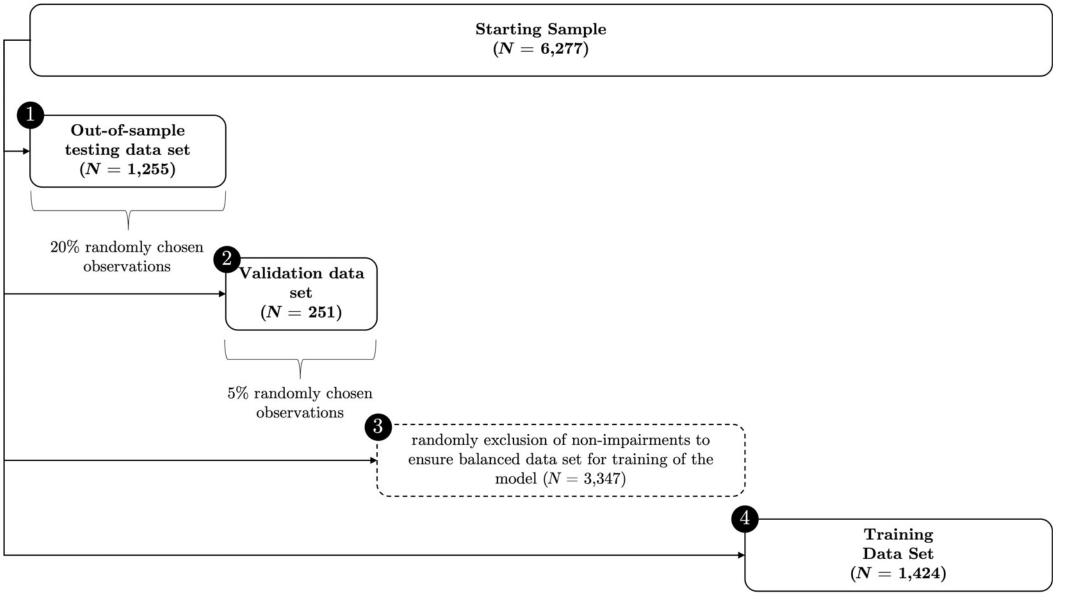

FIGURE 1 Selection of testing, training, and validation samples.

3.3 | Prediction approach

This section describes our prediction approach. First, we outline how we evaluate the performance of our prediction model. Second, we explain the process for selecting the best prediction model during the fine-tuning phase. Third, we introduce our masking process, which reveals the predictive power of specific information within KAMs. Finally, we discuss our approach to investigate the incremental predictive power of goodwill-related KAM disclosures.

3.3.1 | Evaluation metric

To assess the performance of our prediction model, we adopt the area under the receiver operating characteristic curve (area under the curve [AUC]) as our primary evaluation metric (Bao et al., 2020; Bertomeu et al., 2020; N. C. Brown et al., 2020; Chen et al., 2022). The AUC represents the probability that a randomly selected goodwill impairment will be ranked higher by the prediction model than a randomly selected non-impairment (Fawcett, 2006). AUC values from random guessing yield 0.5; therefore, no reasonable prediction model should have an AUC lower than this threshold. We assess whether the AUC of our testing sample prediction models significantly exceeds 0.5 using bootstrap p-values.

3.3.2 | Model selection (RQ1)

We employ the already pre-trained FinBERT language model and fine-tune it for our classification task (Huang et al., 2023). In the fine-tuning process, the language model is trained using a labeled data set comprising KAM text sequences from year t as inputs, with a dichotomous output variable indicating whether the firm impairs goodwill in year $t+1$ or not (*IMPAIR*).

We then use the fine-tuned model to predict goodwill impairments and evaluate its performance on the testing sample.

To find the best prediction model, it is important to select the right hyperparameters for the FinBERT model. To achieve this goal, we employ the hyperparameter optimization framework Optuna developed by Akiba et al. (2019). This approach offers two advantages. First, it is more objective than arbitrary hyperparameter selection. In fact, Optuna allows for automatic optimization of hyperparameters by optimizing an objective function over several trials. Therefore, this algorithm-based optimization approach provides a mathematical solution instead of an arbitrary choice. Second, Optuna has demonstrated superiority over other hyperparameter optimization methods, such as random grid search (Akiba et al., 2019), thereby ensuring that we identify the most effective prediction model with limited computational resources. We employ Optuna with tree-structured Parzen estimators to maximize the AUC of our predictions on the validation sample over 30 trials.[8]

3.3.3 | Masking tests (RQ2)

The prediction model examines whether KAMs are informative about future goodwill impairments. Additionally, we aim to identify which specific information types within KAMs contribute most significantly to their predictive power. To achieve this, we conduct masking tests that yield detailed evidence about the informational content of various aspects reported within KAMs. The core concept of a masking test involves altering certain elements of the input sequences and analyzing how these modifications impact the predictive results of the testing sample (e.g., Siano & Wysocki, 2021). The difference in predictive power between the complete input sequence and the masked input sequence reflects the predictive significance of the masked information type.

We modify or remove specific information types from the KAM sections and compare the AUC to those derived from the complete input sequence. First, we start by removing all *numeric values* in the input sequences as prior research indicates that numbers might represent informative signals within corporate texts (Siano & Wysocki, 2018). Second, we follow Siano and Wysocki (2021) as well as Huang et al. (2023) and randomize all words in the input sequences to examine the importance of context compared to isolated words in KAMs. Third, we sequentially delete each of the six information categories specified by ISA 701. This process entails manually reviewing and annotating all 1,255 KAMs in our testing sample in accordance with the information required by ISA 701. The information categories utilized for masking are as follows: (1) reference to how the matter is addressed in the financial statements (ISA 701 A.40–41), (2) why the matter is considered a KAM (ISA 701 A.42–45), (3) aspects of the auditors' approach to addressing the matter (ISA 701 A.48–49), (4) audit procedures performed (ISA 701 A.50), (5) the outcome of those procedures (ISA 701 A.51), and (6) key observations regarding the matter (ISA 701 A.46). Categories (1) and (2) pertain to the description section of the KAM, while the remaining categories belong to the response section.

3.3.4 | Investigating the incremental predictive power of KAMs (RQ3)

To explore the incremental predictive power of goodwill-related KAMs, we conduct three analyses. First, we examine the semantic similarity between goodwill-related KAMs and corresponding firm

[8]After testing different number of trials, we find 30 trials to be a reasonable compromise between computational resources and model performance. In our case, the hyperparameter optimization process with 30 trials took approximately 6 h for one KAM section. Supplementary Appendix S1 describes the search space for our hyperparameter optimization process in more detail.

disclosures.[9] We manually collect the annual reports for each firm in our testing sample and compute the semantic similarity between the KAM and the firm's disclosures. Given that the response section contains information that—by definition—can only be reported by the auditor, we focus our analysis on the description section of the KAM. We condense the annual report text—without the audit report—to identify potential goodwill-related disclosures by selecting 10 sentences before and 10 sentences after the keyword "goodwill."[10] We then calculate cosine similarity using FinBERT word embeddings for each KAM description sentence and each potential goodwill-related sentence in the firms' annual report. The cosine similarity score ranges from -1 to $+1$, with $+1$ denoting identical semantic similarity. For each KAM sentence, we identify the most similar sentence in the firms' annual report based on the highest similarity score. Finally, we remove sentences exceeding different similarity thresholds and analyze the predictive power of the remaining sentences.

In our second analysis, we assess the incremental predictive power of goodwill-related KAMs beyond goodwill-related firm disclosures.[11] We manually collect the remaining annual reports from firms in our training and validation samples to train a FinBERT model for predicting future impairments, using goodwill-related firm disclosures identified via the same method as in the first analysis. We then evaluate the incremental predictive power of KAMs by predicting impairment probabilities using the KAM text sequences and firm disclosures.

In our third analysis, we investigate the incremental predictive power of goodwill-related KAMs beyond important firm-level variables used in prior literature to explain future goodwill impairments. We employ a nonlinear gradient-boosting machine learning approach to model future impairment probabilities. A literature review identifies several important firm-level variables (e.g., Amel-Zadeh et al., 2020): firm size (*SIZE*), stock returns (*RETURN*), leverage (*LEVERAGE*), firm complexity (*SEGMENTS*), size of goodwill (*GW_RATIO*), impairment history (*IMPAIR_HIST*), firm profitability (*ROA*), growth opportunities (*MTB*), and managerial incentives (*SMOOTH*, *BATH*, *CEO_COMP*). The Appendix describes these variables. We use the firm-level determinants to train a nonlinear model for predicting impairment probabilities, and then compare predictions derived from the goodwill-related KAM text sequences and firm-level determinants to assess KAMs' incremental predictive power. Finally, we evaluate the incremental predictive power of KAMs beyond goodwill-related firm disclosures and the identified firm-level determinants of goodwill impairment.

4 | RESULTS

4.1 | Descriptive statistics

Table 2 presents the descriptive statistics for the KAM text characteristics and goodwill impairments. Panel A reports statistics on the words, sentences, and tokens within the two KAM sections analyzed in our samples. Generally, the response section contains a higher average of words and tokens, but fewer sentences than the description section. In the testing sample, the average description section comprises 166.3 words and 6.6 sentences, while the response section averages 192.7 words and 6.0 sentences. The corresponding averages for tokens are 204.4 for the description section and 223.2 for the response section. We find no statistically

[9]Our approach is similar to Burke et al. (2023), who examine the similarity between CAM disclosures and 10-K footnote language. However, because European firms' notes lack the standardized structure of 10-K footnotes, we first identify goodwill-related disclosures using a self-constructed algorithm.

[10]This step must be performed due to computational constraints.

[11]Our analyses distinguish between textual similarity and statistical modeling: the former identifies and removes overlapping narrative content, while the latter controls for firm disclosures in a predictive framework.

TABLE 2 Descriptive statistics.

Panel A: KAM textual characteristics

	Training and validation sample ($N = 1{,}675$)					Testing sample ($N = 1{,}255$)					Diff.
	Mean	**Min**	**Median**	**Max**	**SD**	**Mean**	**Min**	**Median**	**Max**	**SD**	***t*-stat**
KAM—Description											
Words	168.592	13	150	768	83.337	166.322	19	151	687	83.914	0.728
Sentences	6.643	0	6	34	3.326	6.585	1	6	21	3.279	0.471
Tokens	208.032	18	189	512	99.608	204.371	23	184	512	100.426	0.981
KAM—Response											
Words	193.958	17	182	832	87.287	192.725	18	180	830	87.154	0.378
Sentences	6.124	0	6	25	3.600	5.998	0	6	24	3.474	0.950
Tokens	225.170	20	214	512	94.740	223.210	21	211	512	93.855	0.556

Panel B: Summary statistics of goodwill impairments

	Training and validation sample ($N = 1{,}675$)					Testing sample ($N = 1{,}255$)					Diff.
	Mean	**Min**	**Median**	**Max**	**SD**	**Mean**	**Min**	**Median**	**Max**	**SD**	***t*-stat**
IMPAIR	0.448	0	0	1	0.497	0.168	0	0	1	0.374	16.690***
IMPAIR_SIZE	0.011	0	0	0.535	0.041	0.004	0	0	0.455	0.024	5.422***

Note: This table presents descriptive statistics for our samples. Panel A provides the number of words, sentences, and tokens across KAM sections and the training and validation sample compared to the testing sample. Panel B reports descriptive statistics on the occurrence of impairments (*IMPAIR*) and the magnitude of impairments (*IMPAIR_SIZE*) within our samples.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

significant differences between the textual characteristics of the fine-tuning and testing samples, indicating appropriate prerequisites for our model selection procedures.

Panel B outlines the summary statistics for goodwill impairments in our samples. On average, goodwill impairment occurs in 16.8% of our testing sample observations, aligning with findings from prior studies (e.g., Glaum et al., 2018; Kabir & Rahman, 2016). The average impairment loss is 0.4% of total assets, with a maximum observed loss of 45.5% (*IMPAIR_SIZE*). As detailed in Section 3, we ensure equal representation of impairments and non-impairments in our training sample, resulting in statistically significant differences regarding *IMPAIR* between the fine-tuning and testing data sets.

4.2 | RQ1: Are goodwill-related KAMs informative for future goodwill impairments?

Table 3 presents our main findings. Columns 1–3 report the results of a logistic regression model with *IMPAIR* as the dependent variable. The independent variables of interest include the future impairment probabilities modeled with our FinBERT prediction models for the full KAM text sequence (*KAM_IMP_PROB*), the description section (*DESC_IMP_PROB*), and the response section (*RESP_IMP_PROB*), respectively. The AUCs for the full KAM text, the description section, and the response section range from 79.1% to 82.7%. This reflects a predictive gain of 29.1–32.7 percentage points over random guessing, as illustrated in Figure 2. Importantly, these reported AUCs are significantly greater than 50%—the AUC of a random classifier—indicating that the predictive power of KAMs is unlikely to be a random outcome.

Our additional evaluation metrics further validate the predictive power of KAMs, with Recall scores ranging from 76.8% to 80.6% and Pseudo R^2 ranging from 16.0% to 22.5%. These results imply that the FinBERT algorithm can extract meaningful signals from KAMs, which supports the notion that KAMs are informative for areas of significant judgment in financial statements. The AUCs are comparable to findings from recent studies predicting future earnings (Chen et al., 2022), financial misreporting (N. C. Brown et al., 2020), and fraud (Bao et al., 2020), utilizing multiple firm-level variables. This indicates that KAMs alone possess substantial predictive power for future goodwill impairments, as we disregard all other potential firm-level characteristics from our prediction models.

The lower section of Table 3 shows bootstrap tests for the differences in the reported AUCs. Although no statistically significant differences in predictive power exist between the description and response sections, the AUC for the full KAM text is significantly higher than that of the description section, with a statistically significant difference at the 1% level. This suggests that the description section has predictive power for future goodwill impairment that is comparable to that of the response section. Thus, users can derive meaningful information regarding future goodwill impairments from both KAM sections.

To enhance the interpretability of our findings and provide insights for human readers on the factors that lead the FinBERT prediction model to identify future impairments, we analyze the textual features and content of correctly classified KAMs. This detailed analysis, presented in supplementary Appendix S3, reveals that goodwill-related KAMs associated with future impairments are significantly longer, exhibit more negative sentiment, and are harder to read. Regarding content, these KAMs tend to emphasize the underlying risks and uncertainties related to valuation parameters and impairment testing. Additionally, the audit procedures described in these KAMs often focus on sensitivity analyses, suggesting a narrower margin between the recoverable amount and the corresponding book value of goodwill. Furthermore, auditors refrain from confirming that managerial estimates fall within a reasonable range, unlike their non-impairment counterparts. This

TABLE 3 Main prediction results.

d =	(1) *IMPAIR*	(2) *IMPAIR*	(3) *IMPAIR*
KAM_IMP_PROB	3.705***		
	(0.274)		
DESC_IMP_PROB		2.420***	
		(0.201)	
RESP_IMP_PROB			2.733***
			(0.232)
Constant	−3.693***	−2.930***	−3.180***
	(0.210)	(0.163)	(0.186)
Observations	1,255	1,255	1,255
AUC	**0.827*****	**0.798*****	**0.791*****
Recall	0.768***	0.773***	0.806***
Pseudo R^2	0.225	0.160	0.162
Bootstrap test			
AUC < Column 1		<0.01***	<0.01***
AUC > Column 3		0.340	

Note: This table presents the results of our primary prediction models for the testing sample. Columns 1–3 display the outcomes of a logistic regression with the occurrence of a goodwill impairment in year $t + 1$ as the dependent variable (*IMPAIR*). The independent variables of interest are the probabilities of goodwill impairment, which are predicted using our FinBERT models. These models forecast goodwill impairments for firm i in year $t + 1$, incorporating the full KAMs, the description section, and the response section for firm i in year t. The prediction model employing the description (response) section as input is fine-tuned on the training sample following a 30-trial hyperparameter optimization process as outlined by Akiba et al. (2019). The fine-tuning of the description (response) section yielded a learning rate of 5.41×10^{-5} (8.06×10^{-5}), a weight decay of 1.87×10^{-3} (9.22×10^{-5}), a batch size of 8 (8), a warmup ratio of 0.06 (0.12) and 5 (5) epochs. The AUC values for the description and response sections are computed using the predicted impairment probabilities from the testing sample. The AUC for the full KAM is derived from the average probabilities of the description and response sections within the testing sample. Recall is calculated by dividing the number of correctly classified impairment observations by the total number of impairment observations, considering an observation to be classified as an impairment if the prediction probability exceeds 50%. We evaluate the statistical significance of the evaluation scores by determining whether the statistic exceeds 0.50. This involves conducting one-sample t-tests that compare the evaluation scores of our prediction model against scores generated from simulated random data, using a bootstrap method with 10,000 replications. Additionally, bootstrap tests are performed to examine differences in AUC values among the prediction models, with Wald tests based on 10,000 bootstrap iterations.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

suggests that the FinBERT model leverages these specific textual and content features to effectively distinguish between impairment and non-impairment cases. Collectively, Table 3 shows that the content within goodwill-related KAMs is predictive of future goodwill impairments, supporting the notion that KAMs are informative for financial statement areas involving significant judgment.

4.3 | RQ2: Which type of information within goodwill-related KAMs possesses the highest predictive power?

In this section, we examine how the predictive power of our main results in Table 3 changes when specific aspects of the KAM input sequences are deleted or altered. Table 4 presents the results of these masking tests for the description and response sections. Panel A indicates that deleting all numbers from the description section results in a significant 6.26% decrease in AUC at the 1% level, suggesting that the FinBERT algorithm extracts meaningful signals from these numbers. Randomizing all words decreases the AUC by 17.92%, which is also significant at the 1% level, underscoring the critical role of contextual information. To gain further insights into the content driving the predictive power of the description section, we delete all KAM text passages that elaborate on how

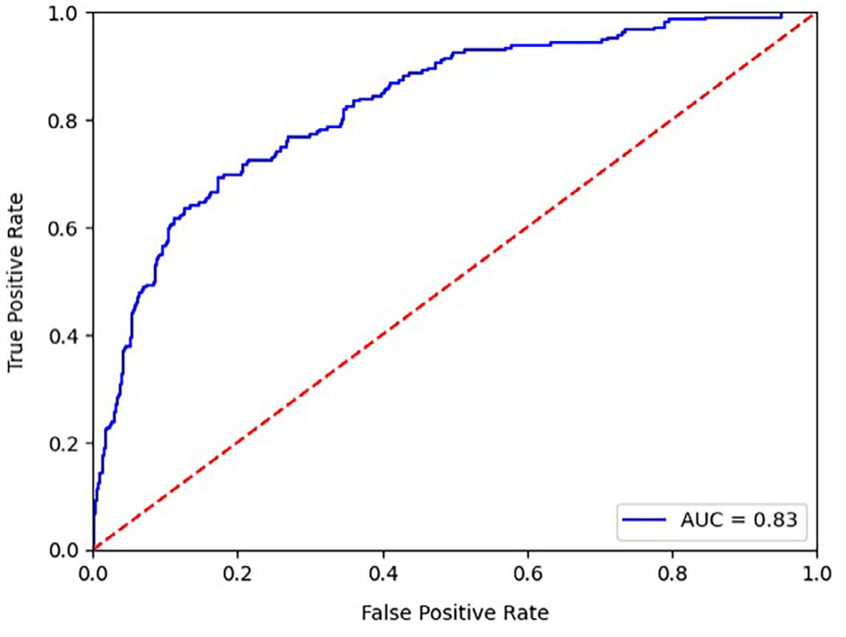

FIGURE 2 Receiver operating characteristic curve for the FinBERT prediction model in Table 3, Column 1.

firms' management addressed goodwill accounting in the financial statements (ISA 701 A.40–41). This deletion results in a statistically significant decrease of 22.93% in predictive power, lowering the AUC to 61.5%. Additionally, when we eliminate all text passages that justify why the auditor considered goodwill as a KAM, the AUC decreases to 71.1%, though still statistically significant. These findings indicate that the predictive power of the description section is predominantly influenced by passages detailing how firms account for goodwill in their financial statements.

In Panel B, we present the results of our masking tests concerning the response section. Unlike our observations in Panel A, we observe no statistically significant change in the AUC when all reported numbers are deleted. However, randomizing the input sequence results in a significant decrease in AUC of 22.88% compared to the full input sequence. When we remove four text passages containing disclosures recommended by ISA 701, we find no statistically significant change in the predictive power of the response section with respect to the outcomes of audit procedures or key auditor observations. In contrast, there is a statistically significant decrease in predictive power when we delete text passages that elaborate on the auditors' approach and the audit procedures performed. Specifically, removing these passages leads to a 24.27% decrease in AUC, compared to a mere 7.08% decrease for passages detailing the auditors' approach. These results indicate that the predictive power of the response section is primarily driven by passages elaborating on the performed audit procedures, highlighting their importance in the overall assessment. Collectively, our masking tests suggest that the predictive power of KAMs is primarily influenced by text passages that provide insights into how the firm and its auditor exercised judgment regarding the accounting and auditing of goodwill.

4.4 | RQ3: Are goodwill-related KAMs incrementally informative of future goodwill impairments?

Table 5 presents our analysis of the semantic similarity between KAMs and the clients' financial statements. Given the absence of an established similarity score threshold, we first provide

TABLE 4 Analysis of KAMs' textual features and contents.

Panel A: Description section of KAM			
Masking test	**AUC**	**Diff.**	**Diff. (%)**
Benchmark (full input sequence)	0.798***		
Deletion of all numbers	0.746***	−0.050***	−6.26
Randomize text sequence	0.655***	−0.143***	−17.92
Deletion of text relating to how the firm addressed goodwill accounting (ISA 701 A.40–41)	0.615***	−0.183***	−22.93
Deletion of text relating to why the matter is considered a KAM (ISA 701 A.42–45)	0.711***	−0.087***	−10.90

Panel B: Response section of KAM			
Masking test	**AUC**	**Diff.**	**Diff. (%)**
Benchmark (full input sequence)	0.791***		
Deletion of all numbers	0.795***	0.004	0.51
Randomize text sequence	0.610***	−0.181***	−22.88
Deletion of text relating to aspects of the auditor's approach (ISA 701 A.48–49)	0.735***	−0.056**	−7.08
Deletion of text relating to performed audit procedures (ISA 701 A.50)	0.599***	−0.192***	−24.27
Deletion of text relating to outcomes of performed audit procedures (ISA 701 A.51)	0.789***	−0.002	−0.25
Deletion of text relating to key observations (ISA 701 A.46)	0.780***	−0.011	−1.39

Note: This table presents the AUC values from our masking tests for the description section in Panel A and the response section in Panel B. The benchmark AUC, representing the full input sequence for both sections, is detailed in Table 3. We assess the statistical significance of the AUCs by determining whether they exceed 0.50, the AUC for a random classifier. To establish significance levels, we perform one-sample t-tests that compare the AUCs of our prediction model with those generated from simulated random data, utilizing 10,000 bootstrap replications. We further evaluate the predictive power of each masking procedure through nonparametric Wald tests, which analyze the differences in AUCs relative to the benchmark based on 10,000 bootstrap iterations.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

illustrative examples of potential thresholds in Panel A of Table 5. As anticipated, these examples demonstrate a decrease in semantic similarity with lower scores. We classify sentences with a similarity score above 0.95 as exhibiting "very high similarity," while those above 0.80 are classified as "low similarity", with gradations in between decreasing in increments of five percentage points.[12] Descriptive statistics for each similarity score threshold are reported in Panel B of Table 5. On average, 7.54% (= 0.483/(5.921 + 0.483) × 100) of KAM description sentences exceed a high similarity score of 0.90.

Panel C of Table 5 reports the results of our masking test for each similarity score threshold.[13] Deleting all sentences with very high similarity yields no significant impact on the AUC, indicating that these sentences do not drive the predictive power of the KAM description section. In contrast, we observe a statistically significant decrease in AUC for the remaining similarity score thresholds, ranging from approximately 4% to 28%. The highest decrease occurs when deleting all sentences with a similarity score above 0.80, suggesting that semantic similarity is not the primary driver of the predictive power of the description section.

[12]Our approach results in generally high similarity scores, as we align each KAM sentence with the most similar sentence from the corresponding annual report.

[13]As we lose some observations for missing annual reports, we replicate the prediction model for the full description sequence on the reduced sample size.

TABLE 5 The incremental predictive power of KAMs beyond goodwill-related disclosures in the annual report.

Panel A: Illustrative examples for similar sentences in each similarity score threshold

Similarity score threshold	Sentence within KAM	Most similar sentence in the firms' annual report	Similarity score
>0.95 (very high)	"An impairment loss is recognized if the carrying amount of the cash-generating unit exceeds its recoverable amount."—Tecan Group Ltd. 2020 annual report (Tecan Group Ltd., 2021, p. 169)	"An impairment loss is only recognized if the carrying amount of the cash-generating unit exceeds its recoverable amount."—Tecan Group Ltd. 2020 annual report (Tecan Group Ltd., 2021, p. 153)	0.994
>0.90 (high)	"As described in note 10 to the consolidated financial statements, goodwill, which arose on the acquisition of the wholly owned subsidiary, Scancell Limited, is considered to have an indefinite useful life."—Scancell Holdings plc, 2018 annual report (Scancell Holdings plc, 2018, p. 16)	"The goodwill arose on the acquisition of the wholly owned subsidiary Company, Scancell Limited and is considered to have an indefinite life."—Scancell Holdings plc, 2018 annual report (Scancell Holdings plc, 2018, p. 29)	0.932
>0.85 (medium)	"Determining the recoverable amount of these assets and any impairment losses to be recognised is a key audit matter, given the importance of estimates and the level of judgment required by management regarding the operational performance and future traffic assumptions, long-term growth rates and discount rates used, and the sensitivity of their measurement to changes in certain assumptions."—Vinci SA 2020 annual report (Vinci SA, 2021, p. 355)	"The assumptions and estimates made to determine the recoverable amount of goodwill, intangible assets and property, plant and equipment relate in particular to the assessment of market prospects needed to estimate the cash flow, and the discount rates adopted."—Vinci SA 2020 annual report (Vinci SA, 2021, p. 293)	0.887
>0.80 (low)	"Goodwill relating to the ingenie cash generating unit ("ingenie") is significant and at risk of impairment should the business not generate sufficient future economic benefits."—Watchstone Group plc 2017 annual report (Watchstone Group plc, 2018 p. 25)	"As part of the strategic restructuring of the Group, two business operations within the then Hubio business segment were closed during the year, namely the North American telematics operation and Hubio's IT development operation in Dundee."—Watchstone Group plc 2017 annual report (Watchstone Group plc, 2018, p. 7)	0.830

Panel B: Descriptive statistics for each similarity score threshold

Similarity score threshold	Average number of similar sentences	Average number of remaining dissimilar sentences
>0.95	0.100	6.304
>0.90	0.483	5.921

(Continues)

TABLE 5 (Continued)

Panel B: Descriptive statistics for each similarity score threshold

Similarity score threshold	Average number of similar sentences	Average number of remaining dissimilar sentences
>0.85	2.412	3.992
>0.80	4.667	1.737

Panel C: Masking tests for each similarity score threshold

Masking test	AUC	Diff.	Diff. (%)
Benchmark (full description section)	0.814***		
Delete sentences with similarity score >0.95	0.813***	−0.001	−0.00
Delete sentences with similarity score >0.90	0.782***	−0.032***	−3.93
Delete sentences with similarity score >0.85	0.703***	−0.111***	−13.64
Delete sentences with similarity score >0.80	0.587***	−0.227***	−27.89

Note: Panel A presents illustrative examples for each similarity score threshold. Panel B provides descriptive statistics corresponding to these thresholds. Panel C details the masking tests associated with each similarity score threshold. The number of observations is contingent upon the availability of annual reports for the firms in the testing sample. We assess the statistical significance of the AUC by determining whether it exceeds 0.50, which represents the AUC of a random classifier. To evaluate significance, we perform one-sample *t*-tests comparing the AUCs of our prediction model with those generated from simulated random data, employing a bootstrap method with 10,000 replications. Additionally, we analyze the predictive power of each masking procedure using nonparametric Wald tests to compare the AUCs against the benchmark. These Wald tests also utilize 10,000 bootstrap iterations.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

Table 6 presents our analysis testing the incremental predictive power of KAMs beyond goodwill-related firm disclosures and important firm-level variables. Given that our sample size is contingent upon available firm-level data and annual reports, we first replicate our findings from Table 3 on a reduced sample of 966 observations, as shown in Column 1 of Table 6. The resulting AUC of 82.6% is comparable to our main results with the full sample of 1,255 observations. Column 2 presents the results of a logistic regression model with *IMPAIR* as the dependent variable, utilizing future impairment probabilities derived from goodwill-related disclosures within firms' financial statements (*DISC_IMP_PROB*) as the independent variable of interest. The model achieves an AUC score of 70.5%, which is significantly lower than that of the model in Column 1.

Columns 3 and 4 detail the findings of logistic regression models including important firm-level variables to predict future impairments. Column 3 employs a linear model, while Column 4 utilizes a gradient-boosting machine learning algorithm, modeling impairment probabilities based on a combination of firm-level determinants related to goodwill impairment (*FUNDA_IMP_PROB*). In line with prior literature, we find impairment history (*IMPAIR_HIST*), future growth opportunities (*MTB*), and goodwill ratio (*GW_RATIO*) to be significantly associated with future impairments (e.g., Glaum et al., 2018). The Pseudo R^2 ranges from 0.08 to 0.10 and is slightly below those reported in other studies (e.g., Glaum et al., 2018; Han et al., 2021; Ramanna & Watts, 2012).[14] Regarding AUC, the predictive power of firm-level determinants remains significantly lower than that of goodwill-related KAMs, with the AUC for the nonlinear model in Column 4 yielding 73.1%.

In Column 5, we combine *DISC_IMP_PROB* and *FUNDA_IMP_PROB* to establish our benchmark model, which shows an increased AUC compared to the models in Columns 2 and 4. Column 6 integrates all impairment probabilities into a single model to evaluate the

[14]Our sample might not be comparable to samples of prior studies, as we focus on a subset of firms with goodwill-related KAMs. Thus, the determinants of goodwill impairments might also differ for this subset of firms.

TABLE 6 The incremental predictive power of KAMs beyond firm-level disclosures and determinants of goodwill impairment.

	KAM only	Firm-level disclosures	Firm-level determinants		Firm-level disclosures and determinants	Incremental predictive power
	(1)	(2)	(3)	(4)	(5)	(6)
d =	*IMPAIR*	*IMPAIR*	*IMPAIR*	*IMPAIR*	*IMPAIR*	*IMPAIR*
KAM_IMP_PROB	3.753***					3.080***
	(0.313)					(0.410)
DISC_IMP_PROB		2.644***			2.212***	1.210***
		(0.330)			(0.343)	(0.374)
FUNDA_IMP_PROB				7.252***	6.316***	4.492***
				(0.805)	(0.830)	(0.916)
SIZE			0.147			
			(0.114)			
RETURN			−0.194			
			(0.128)			
LEVERAGE			0.024			
			(0.087)			
SEGMENTS			−0.075			
			(0.099)			
GW_RATIO			−0.227**			
			(0.112)			
IMPAIR_HIST			0.451***			
			(0.082)			
ROA			0.126			
			(0.132)			
MTB			−0.766***			
			(0.228)			
SMOOTH			0.114			
			(0.087)			
BATH			−0.296			
			(0.431)			
CEO_COMP			0.128			
			(0.096)			
Constant	−3.697***	−2.904***	−1.804***	−4.193***	−4.964***	−5.533***
	(0.240)	(0.208)	(0.111)	(0.323)	(0.365)	(0.410)
Observations	966	966	966	966	966	966
AUC	0.826***	0.705***	0.707***	0.731***	0.767***	0.845***
Pseudo R^2	0.229	0.081	0.083	0.101	0.151	0.276
Bootstrap test						
AUC < Column 1		<0.01***	<0.01***	<0.01***	<0.01***	
AUC > Column 1						<0.01***
AUC > Column 5						<0.01***

Note: This table presents the results of logistic regressions with the occurrence of goodwill impairment in $t + 1$ as the dependent variable (*IMPAIR*) across all columns. The number of observations is contingent upon the availability of firm-level control variable data and firms' annual reports. In Column 1, the independent variable of interest, *KAM_IMP_PROB*, represents impairment probabilities modeled using the complete KAM text sequence with FinBERT. In Column 2, the variable *DISC_IMP_PROB* denotes impairment probabilities derived from goodwill-related sentences within firms' annual reports, utilizing a FinBERT prediction model. This model was fine-tuned on a training sample following a 30-trial hyperparameter optimization run as outlined by Akiba et al. (2019). The tuning process resulted in a learning rate of 6.05×10^{-5}, a weight decay of 2.82×10^{-6}, a batch size of 16, a warmup ratio of 0.09, and 4 epochs. In Column 4, the variable *FUNDA_IMP_PROB* is modeled using a gradient-boosting algorithm based on the firm-level variables presented in Column 3. To optimize this model, we again employed a 30-trial hyperparameter optimization run in accordance with Akiba et al. (2019), achieving a learning rate of 0.0036, a maximum depth of 5, 520 estimators, and a subsample ratio of 0.46. We evaluate the statistical significance of the evaluation score AUC by determining whether the statistic is significantly greater than 0.50. To assess significance levels, one-sample t-tests compare the evaluation score of our prediction model against those generated from simulated random data bootstrapped with 10,000 replications. Additionally, bootstrap tests evaluate differences in AUCs among the prediction models, with Wald tests conducted over 10,000 bootstrap iterations. A detailed definition of the variables is provided in the Appendix.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

incremental predictive power of KAMs.[15] The coefficient for *KAM_IMP_PROB* is smaller than that in the univariate model presented in Column 1, indicating that some explanatory power of the KAM variables is absorbed by the firm-level disclosures and other determinants. Nevertheless, the KAM variable remains statistically significant at the 1% level with a positive coefficient. Compared to the benchmark model in Column 5, we observe an increase in predictive power of 7.8 percentage points. Utilizing bootstrap *p*-values with 10,000 iterations, we confirm that this increase in predictive power is statistically significant.

Collectively, these findings suggest that a goodwill prediction model benefits from the informational content of KAM disclosures, indicating that goodwill-related KAMs are also predictive beyond goodwill-related disclosures within firms' financial statements and firm-level determinants of future goodwill impairments.

5 | ADDITIONAL ANALYSES

5.1 | Are KAM-predicted impairment probabilities associated with market reactions?

To further explore the informational content of goodwill-related KAMs, we investigate market reactions to KAM-predicted impairment probabilities. In particular, we focus on short-window price reactions surrounding the auditor's report publication date.[16] Given that higher impairment probabilities are likely perceived as negative news by investors (e.g., Amel-Zadeh et al., 2020), we use signed cumulative abnormal returns (*CAR*) to capture the average change in investors' expectations (e.g., Kim & Verrecchia, 1991). If goodwill-related KAMs convey incrementally new information to the market, we expect a negative price reaction.

Table 7 presents the results of OLS regressions with *CAR* as the dependent variable.[17] In Column 1, we find no statistically significant association between KAM-predicted impairment probabilities and price reactions, since the coefficient estimate of *KAM_IMP_PROB* is not statistically significant. This suggests that goodwill-related KAMs, on average, do not provide investors with incrementally new information. One possible explanation could be that in the case of rich information environments—typically characterized by multiple information sources—investors may have already anticipated the higher impairment risk (Lennox et al., 2023). This "anticipation argument" is further supported by the FRC, which proposes that the informational benefits of KAMs may be "particularly important for audited entities where there are fewer sources of other information, including smaller companies" (FRC, 2016, p. 4). Thus, we examine whether price reactions differ between small and large firms. We partition firms based on the sample median of *SIZE*, introducing two indicator variables: *SMALL* and *LARGE*. We then incorporate

[15]Untabulated correlation coefficients between *KAM_IMP_PROB*, *DISC_IMP_PROB*, and *FUNDA_IMP_PROB* are all below |0.50|. Variance inflation factors (VIFs) are also well below the critical value of 10, suggesting that multicollinearity is not a concern when including these three variables in the model.

[16]Unlike in the United States, where EDGAR provides precise filing dates, Europe lacks a centralized, reliable source. Filing dates in commercial databases often reflect earnings announcements instead, making them unreliable (Gutierrez et al., 2018). Following prior European studies (e.g., Ianniello & Galloppo, 2015; Soltani, 2000), we use the audit opinion date as a proxy for the audit report release date. To validate this approach, we manually checked 100 annual reports in our sample where this information was available through press releases and corporate websites. We found that in 84% of cases, the audit opinion date matched the actual publication date, with an average deviation of only 0.39 days. Additionally, re-estimating the regressions from Table 7 using a 3-day window following the audit opinion date yields consistent results.

[17]Building on the regression model from Table 6, we add control variables that might affect investor reactions around the release of the audit report, including earnings surprise (*SURPRISE*), the magnitude of reported extraordinary items (*SPECIAL_ITEMS*), analyst coverage (*FOLLOWING*), and the number of disclosed KAMs (*NUM_KAM*). We also include the impairment probabilities modeled non-linearly using firm-level determinants (*FUNDA_IMP_PROB*) and firm disclosures (*DISC_IMP_PROB*) in our regression models to isolate the incremental effect of KAMs. While not all disclosure channels can be captured, our analysis targets the most prominent sources of information on future goodwill impairments. We also acknowledge potential confounding effects when investigating market reactions to audit reports (e.g., Czerney et al., 2019; Ogneva & Subramanyam, 2007).

TABLE 7 Market reactions of KAM-predicted impairment probability.

	(1) *CAR* [−1,+1]	(2) *CAR* [−1,+1]
KAM_IMP_PROB	−0.005	
	(0.004)	
KAM_IMP_PROB×LARGE		0.014
		(0.008)
KAM_IMP_PROB×SMALL		−0.025***
		(0.005)
SMALL		0.021
		(0.012)
DISC_IMP_PROB	0.041	0.039
	(0.023)	(0.022)
FUNDA_IMP_PROB	−0.024	−0.027
	(0.030)	(0.032)
Observations	937	937
Controls	Yes	Yes
Industry fixed effects	Yes	Yes
Region fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
Highest VIF	3.91	5.60
Adj R^2	0.041	0.049
F-test for difference (*p*-value): *KAM_IMP_PROB×LARGE* = *KAM_IMP_PROB×SMALL*		0.009

Note: This table presents ordinary least squares regressions analyzing short-term price reactions, measured by cumulative abnormal returns (*CAR*), over a 3-day period surrounding the audit opinion date. The number of observations is contingent upon the availability of firm-level control variable data. The control variables include *SIZE*, *RETURN*, *LEVERAGE*, *SEGMENTS*, *GW_RATIO*, *IMPAIR_HIST*, *ROA*, *MTB*, *SMOOTH*, *BATH*, *CEO_COMP*, *NUM_KAM*, *SPECIAL_ITEMS*, *SURPRISE*, *FOLLOWING*, *DISC_IMP_PROB*, and *FUNDA_IMP_PROB*. A detailed definition of the variables is provided in the Appendix.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

interaction terms between *KAM_IMP_PROB* and both *SMALL* and *LARGE* into our model.[18] Column 2 of Table 7 shows a statistically significant negative association between *KAM_IMP_PROB* and *CAR* for small firms (coeff. = −0.025, *p*-value <0.01), while no significant effect is observed for large firms. The difference between the two coefficient estimates is statistically significant (*p*-value <0.01).[19] This finding suggests that the higher impairment risk disclosed in KAMs provides incrementally new information only to investors of small firms, which is likely due to fewer available information sources. In contrast, investors of large firms were likely already aware of the higher impairment risk, given their access to a richer information environment.

To substantiate this "anticipation argument" for large firms, we conduct additional analyses using long-window abnormal buy-and-hold returns (BHARs), as detailed in supplementrary Appendix S4. For instance, we find that KAM-predicted impairment probabilities are indeed

[18]We use a partitioning approach—instead of a base approach—to enable an intuitive interpretation of the effects across small and large firms (e.g., Cen et al., 2025; Darendeli et al., 2022; Holzman et al., 2021; Yip & Tsang, 2007).
[19]This test corresponds to the coefficient estimate of *KAM_IMP_PROB×SMALL* in a base approach regression model, which includes *KAM_IMP_PROB*, SMALL, and *KAM_IMP_PROB×SMALL*.

TABLE 8 Predicting the magnitude of goodwill impairments.

d =	(1) *IMPAIR_SIZE*	(2) *IMPAIR_SIZE*	(3) *IMPAIR_SIZE*
KAM_IMP_SIZE	0.362***		
	(0.017)		
DESC_IMP_SIZE		0.415***	
		(0.041)	
RESP_IMP_SIZE			0.190***
			(0.010)
Constant	−0.040***	−0.050	−0.007
	(0.005)	(0.009)	(0.004)
Observations	1,255	1,255	1,255
Adj R^2	0.255	0.075	0.233
Bootstrap test			
Adj R^2 < Col (1)		<0.01***	
Adj R^2 < Col (1)			0.993
Adj R^2 > Col (2)			<0.01***

Note: This table presents prediction results for the magnitude of goodwill impairments. Columns 1–3 display the outcomes of an ordinary least squares regression, with the magnitude of goodwill impairment in year $t + 1$ (*IMPAIR_SIZE*) as the dependent variable. The independent variables of interest are the predicted magnitudes of goodwill impairment, modeled using FinBERT for firm i in year $t + 1$. This modeling incorporates the full KAMs, the description section, and the response section of firm i in year t. The prediction model, optimized through a 30-trial hyperparameter tuning process based on the framework outlined by Akiba et al. (2019), employs the description (response) section as input on the training sample with a learning rate of 3.60×10^{-5} (5.63×10^{-5}), a weight decay of 6.95×10^{-5} (2.64×10^{-6}), a batch size of 8 (16), a warmup ratio of 0.02 (0.28) and 5 (5) epochs. Bootstrap tests are conducted to evaluate the differences in the adjusted R^2 for the prediction models. The Wald tests are performed based on 10,000 bootstrap iterations.
*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

already reflected in returns for large firms up to 12 months prior to the KAM publication, providing empirical support for our "anticipation argument." Collectively, our CAR and BHAR analyses suggest that goodwill-related KAMs are value-relevant for firms of all sizes, but provide incrementally new information, at the time of their publication, only to investors of small firms.

5.2 | Are KAMs predictive of the magnitude of goodwill impairments?

While our previous analyses focus on the predictive power of goodwill-related KAMs on the incidence of future goodwill impairments, we now investigate whether they can also predict their magnitude. As the pre-defined FinBERT network architecture is designed for classification tasks, we follow the approach outlined in Siano (2022) and add a linear layer to allow for continuous predictions. For hyperparameter optimization, we employ the same method as used in our primary prediction models.

Table 8 presents the results of OLS regressions, with impairment size (*IMPAIR_SIZE*) as the dependent variable and KAM-predicted impairment amounts as independent variables. Our findings indicate that each KAM section is predictive of the magnitude of future goodwill impairments, as reflected by adjusted R^2 values ranging from 0.08 to 0.26.[20] Notably, significant

[20]We note that the Adj R^2 is comparable to, or higher than, that of previous studies (e.g., Abdul Majid, 2015; Han et al., 2021; Z. Li et al., 2011). Please note that in a continuous prediction setting, our evaluation metrics AUC and Recall are not adequate.

TABLE 9 The informational content of CAMs and other KAM topics.

	CAM—Impairment			KAM—Business combinations			KAM—Going concern		
$d =$	(1) *IMPAIR*	(2) *IMPAIR*	(3) *IMPAIR*	(4) *RESTR*	(5) *RESTR*	(6) *RESTR*	(7) *GC_OPINION*	(8) *GC_OPINION*	(9) *GC_OPINION*
Full CAM	2.906***								
	(0.455)								
Full KAM				2.258***			2.973***		
				(0.342)			(0.343)		
Description		2.537***			1.471***			2.753***	
		(0.388)			(0.218)			(0.325)	
Response			1.765***			1.580***			1.918***
			(0.368)			(0.280)			(0.257)
Constant	−1.994***	−1.753***	−1.387***	−2.781***	−2.309***	−2.354***	−2.552***	−2.380***	−2.057***
	(0.283)	(0.245)	(0.238)	(0.263)	(0.266)	(0.222)	(0.237)	(0.220)	(0.200)
Observations	391	391	391	556	556	556	436	436	436
AUC	0.694***	0.691***	0.636***	0.717***	0.695***	0.683***	0.804***	0.787***	0.738***
Recall	0.752***	0.726***	0.739***	0.755***	0.764***	0.736***	0.735***	0.699***	0.726***
Pseudo R^2	0.090	0.093	0.047	0.094	0.063	0.066	0.185	0.167	0.127

Note: Columns 1–3 present the results of a logistic regression with the occurrence of goodwill impairment in year $t+1$ as the dependent variable (*IMPAIR*). The independent variables of interest are the probabilities of goodwill impairment, modeled using FinBERT models to predict impairments for firm i in year $t+1$. This modeling employs the full CAM, along with the description and response sections of goodwill-related CAMs for firm i in year t. Following a 30-trial hyperparameter optimization as outlined by Akiba et al. (2019), the prediction model, using the description (response) section as input, is fine-tuned on the training sample with a learning rate of 6.66×10^{-5} (1.05×10^{-5}), a weight decay of 3.11×10^{-4} (7.58×10^{-3}), a batch size of 16 (4), a warmup ratio of 0.15 (0.09), and 3 (5) epochs. Starting with 1,957 observations for goodwill-related CAMs, 391 observations (20%) are designated as the testing sample, 78 observations (5%) as the validation sample, and a balanced sample of 1,034 observations is used for model training. Columns 4–6 report the results of a logistic regression with the occurrence of restructuring expenses in year $t+1$ as the dependent variable (*RESTR*). The independent variables of interest are the probabilities for restructuring expenses, also modeled using FinBERT models. This model predicts restructuring expenses for firm i in year $t+1$ with the full KAM, as well as the description and response sections of Business Combination KAM types for firm i in year t. As a result of a 30-trial hyperparameter optimization following the framework developed by Akiba et al. (2019), the prediction model, using the description (response) section as input, is fine-tuned on the training sample with a learning rate of 4.72×10^{-5} (2.12×10^{-5}), a weight decay of 4.67×10^{-4} (1.12×10^{-3}), a batch size of 8 (4), a warmup ratio of 0.02 (0.18), and 5 (4) epochs. Initially, 2,782 observations related to business combinations were used, with 556 observations (20%) as the testing sample, 111 observations (5%) as the validation sample, and a balanced sample of 910 observations for training. Columns 7–9 display the results of the logistic regression with the occurrence of a going-concern opinion in year $t+1$ as the dependent variable (*GC_OPINION*). The independent variables of interest are the probabilities of issuing a going-concern opinion, modeled with FinBERT models to predict this opinion for firm i in year $t+1$ using the full KAM and the description and response sections of Going-Concern KAM types for firm i in year t. Following a 30-trial hyperparameter optimization according to the framework by Akiba et al. (2019), the prediction model, using the description (response) section as input, is fine-tuned on the training sample with a learning rate of 4.52×10^{-5} (5.15×10^{-5}), a weight decay of 2.48×10^{-6} (8.30×10^{-3}), a batch size of 16 (4), a warmup ratio of 0.09 (0.06), and 4 (5) epochs. With 2,182 observations for going concern–related KAMs, we designate 436 observations (20%) as the testing sample, 87 observations (5%) as the validation sample, and a balanced sample of 798 observations for training. We evaluate the statistical significance of the evaluation scores by assessing whether the statistic is significantly greater than 0.50. To determine significance levels, we calculate one-sample t-tests by comparing the evaluation scores of our prediction model with those generated from simulated random data, bootstrapped with 10,000 replications. Bootstrap tests are also conducted to assess differences in the AUCs of the prediction models, and Wald tests are based on 10,000 bootstrap iterations.

*, **, and *** represent significance levels of 0.10, 0.05, and 0.01, respectively.

differences exist among the KAM sections. The response section accounts for approximately 23% of the variation in future impairment magnitude, exceeding the predictive power of the description section, which explains around 8%. This suggests that auditors' responses regarding performed audit procedures are particularly valuable for predicting the magnitude of future impairments. Collectively, the results in Table 8 support the notion that goodwill-related KAMs are also predictive of the magnitude of future impairments.

5.3 | Are our findings limited to goodwill-related KAMs?

In our final additional analysis, we test whether our findings generalize beyond goodwill-related KAMs. First, we replicate our main prediction results using goodwill-related CAMs data from the US market, as shown in Table 9, Columns 1–3. Despite the constrained data set, our analysis indicates that CAMs are also predictive of future goodwill impairments, with AUCs ranging from 63.6% to 69.4%. While slightly lower than the KAM-based models, this reinforces the idea that the predictive power of KAMs is also applicable in the United States.

Second, we examine whether the predictive power of KAMs extends beyond those related to goodwill. Specifically, we investigate the predictive power of two commonly reported KAM topics associated with significant judgments: business combinations and going concern. For business combinations, we test whether KAMs predict future restructuring expenses that may arise post-merger or acquisition. Columns 4–6 of Table 9 provide evidence that KAMs possess stand-alone predictive power for future restructuring expenses, with AUCs ranging from 68.3% to 71.7%. Regarding going-concern KAMs, we investigate their predictive power for future going-concern opinions issued by the auditor. The results, presented in Columns 7–9 of Table 9, demonstrate that going concern−related KAMs also exhibit stand-alone predictive power, with AUCs ranging from 73.8% to 80.4%. Overall, these results indicate that our main findings are not limited to Europe or goodwill-related KAMs.

6 | ROBUSTNESS

We test the robustness of our main findings in two ways. First, we benchmark FinBERT against traditional "bag-of-words" approaches, using several textual features of KAMs as well as combinations of the 1,000 most frequently used uni-, bi-, and trigrams within KAMs to predict future impairments. As shown in supplementary Appendix S5.1, FinBERT significantly outperforms these alternatives, confirming its suitability for modeling KAM text.

Second, we test the sensitivity of our main results through (1) five-fold cross-validation, (2) focusing exclusively on material impairments as our label, (3) controlling for potential language differences, and (4) utilizing the BERT large language model. Supplementary Appendix S5.2 shows that our main findings hold across varying sample compositions, levels of goodwill materiality, language differences, and the application of a non−domain-specific language model.

7 | CONCLUSION

This study provides empirical evidence on the informational content of KAMs. Using FinBERT—a state-of-the-art NLP deep learning model—we find that the content of goodwill-related KAMs is predictive for future goodwill impairments of a firm. Both individual KAM sections and the full KAM text exhibit stand-alone predictive power that significantly exceeds that of a random classifier, with gains ranging from 29.1 to 32.7 percentage points. Exploring the semantic meaning of goodwill-related KAMs, we find that disclosures about (1) how the firm accounts for goodwill in the financial statements and (2) how the auditor responds with audit procedures possess the highest predictive power. Additionally, we demonstrate that goodwill-related KAMs are incrementally predictive *beyond* (1) goodwill-related disclosures in firms' annual reports and (2) key firm-level determinants. Further analyses of (1) market reactions to KAM-predicted impairment probabilities, (2) the prediction of impairment magnitudes, (3) CAMs, and (4) other KAM topics corroborate our main findings. Overall, our results support the assertion that KAMs hold informational content for financial statement areas characterized by significant management judgment.

Our results have implications for regulators, auditors, and practitioners. From a regulatory perspective, they support the notion that KAMs fulfill one of their intended aims—having informational content. By demonstrating the predictive power of goodwill-related KAMs, we establish their role in enhancing users' understanding of financial statement areas involving significant judgment. Specifically, disclosures regarding how management and auditors exercise their judgment are particularly valuable, suggesting that regulators might consider enhancing such disclosures. From an auditor's perspective, our findings illustrate that KAMs serve as an effective mechanism for informing stakeholders about potential firm risks. Auditors can leverage KAMs to enhance the understanding of complex accounting topics and how they relate to future outcomes. From a practitioners' perspective, our findings indicate the potential of FinBERT as a predictive tool for future negative accounting outcomes. Given that FinBERT, a large language model, emulates human text comprehension, its integration within a machine learning prediction algorithm positions it as a valuable asset for analyzing accounting documents.

Our study is subject to three main limitations. First, the recent adoption of KAM reporting in Europe limits our sample size. While FinBERT handles small samples effectively, we cannot assess predictive power over longer horizons. Second, our masking tests rely on manual annotation of KAM text, which—despite independent validation—remains subjective. Third, due to data constraints, our models focus on firm-wide goodwill impairments and cannot distinguish between impairments at the level of individual cash-generating units.

Despite these limitations, our study contributes to the accounting literature by being the first to provide *direct* evidence of the predictive power of KAMs and to identify the types of information associated with the highest predictive power. This is not only important for regulators and auditors to evaluate audit reporting regulation, but also enhances our general understanding of the informational content of KAMs.

ACKNOWLEDGMENTS

We thank all participants of the 45th EAA Annual Congress in Helsinki, the 2023 AAA Annual Meeting in Denver, the 12th EARNet Symposium in Thessaloniki, and research seminars at San Diego State University 2025, University of Bamberg 2023, and Catholic University of Eichstätt-Ingolstadt 2022. We are especially grateful for the valuable comments and suggestions from Matthew Lyle (editor), two anonymous reviewers, Dennis Ahn, Jochen Bigus, June Cao, Hung Chan, Brigitte Eierle, Mahmoud Elmarzouky (discussant), Nadine Georgiou, James Hansen (discussant), Sven Hartlieb, Jochen Hundsdoerfer, Siegfried Köstlmeier, Valerie Li, Maximilian Nagl, Francesco Mazzi, Javad Rajabalizadeh, Daniel Rösch, Klaus Röder, Klaus Ruhnke, Frank Schiemann, Myles Stern, David Streich, Jeff Wang, and Tonni Xia. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data used in this study stems from LSEG and the Audit Analytics database.

REFERENCES

Abdelfattah, T., Elmahgoub, M., & Elamer, A. A. (2021). Female audit partners and extended audit reporting: UK evidence. *Journal of Business Ethics*, *174*(1), 177–197. https://doi.org/10.1007/s10551-020-04607-0

Abdul Majid, J. (2015). Reporting incentives, ownership concentration by the largest outside shareholder, and reported goodwill impairment losses. *Journal of Contemporary Accounting & Economics*, *11*(3), 199–214. https://doi.org/10.1016/j.jcae.2015.07.002

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623–2631). Association for Computing Machinery. https://doi.org/10.1145/3292500.3330701

Amel-Zadeh, A., Glaum, M., & Sellhorn, T. (2020). Empirical goodwill research: Insights, issues, and implications for standard setting and future research. *European Accounting Review*, *32*(2), 415–446. https://doi.org/10.1080/09638180.2021.1983854

Andreicovici, I., Jeny, A., & Lui, D. (2020). Do firms respond to auditors' red flags? Evidence from the expanded audit report. Working paper. https://doi.org/10.2139/ssrn.3634479

Araci, D. (2019). *FinBERT: Financial sentiment analysis with pre-trained language models*. arXiv. http://arxiv.org/abs/1908.10063

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Beatty, A., & Weber, J. (2006). Accounting discretion in fair value estimates: An examination of SFAS 142 goodwill impairments. *Journal of Accounting Research*, *44*(2), 257–288. https://doi.org/10.1111/j.1475-679X.2006.00200.x

Bédard, J., Gonthier-Besacier, N., & Schatt, A. (2019). Consequences of expanded audit reports: Evidence from the justifications of assessments in France. *Auditing: A Journal of Practice & Theory*, *38*(3), 23–45. https://doi.org/10.2308/ajpt-52339

Bepari, M. K., Mollik, A. T., Nahar, S., & Islam, M. N. (2022). Determinants of accounts level and entity level key audit matters: Further evidence. *Accounting in Europe*, *19*(3), 397–422. https://doi.org/10.1080/17449480.2022.2060753

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2020). Using machine learning to detect misstatements. *Review of Accounting Studies*, *26*(2), 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bingler, J. A., Kraus, M., Leippold, M., & Webersinke, N. (2022). Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters*, *47*, 102776. https://doi.org/10.1016/j.frl.2022.102776

Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *Contemporary Accounting Research*, *40*(2), 765–805. https://doi.org/10.1111/1911-3846.12825

Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review*, *95*(2), 31–60. https://doi.org/10.2308/accr-52507

Boolaky, P. K., & Quick, R. (2016). Bank directors' perceptions of expanded auditor's reports. *International Journal of Auditing*, *20*(2), 158–174. https://doi.org/10.1111/ijau.12063

Brasel, K., Doxey, M. M., Grenier, J. H., & Reffett, A. (2016). Risk disclosure preceding negative outcomes: The effects of reporting critical audit matters on judgments of auditor liability. *The Accounting Review*, *91*(5), 1345–1362. https://doi.org/10.2308/accr-51380

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using *topic* to detect financial misreporting. *Journal of Accounting Research*, *58*(1), 237–291. https://doi.org/10.1111/1475-679X.12294

Brown, S. V., Hinson, L. A., & Tucker, J. W. (2024). Financial statement adequacy and firms' MD&A disclosures. *Contemporary Accounting Research*, *41*(1), 126–162. https://doi.org/10.1111/1911-3846.12919

Burke, J. J., Hoitash, R., Hoitash, U., & Xiao, S. (2023). The disclosure and consequences of US critical audit matters. *The Accounting Review*, *98*(2), 59–95. https://doi.org/10.2308/TAR-2021-0013

Camacho-Miñano, M.-M., Muñoz-Izquierdo, N., Pincus, M., & Wellmeyer, P. (2024). Are key audit matter disclosures useful in assessing the financial distress level of a client firm? *The British Accounting Review*, *56*(2), Article 101200. https://doi.org/10.1016/j.bar.2023.101200

Cen, X., Li, N., Tang, C., & Wang, J. (2025). Preference dynamics and risk-taking incentives. *Journal of Accounting and Economics*, *79*(2–3), Article 101739. https://doi.org/10.1016/j.jacceco.2024.101739

Chen, X., Cho, Y. H., Dou, Y., & Lev, B. (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research*, *60*(2), 467–515.

Christensen, B. E., Glover, S. M., & Wolfe, C. J. (2014). Do critical audit matter paragraphs in the audit report change nonprofessional investors' decision to invest? *Auditing: A Journal of Practice & Theory*, *33*(4), 71–93. https://doi.org/10.2308/ajpt-50793

Church, B. K., Davis, S. M., & McCracken, S. A. (2008). The auditor's reporting model: A literature overview and research synthesis. *Accounting Horizons*, *22*(1), 69–90. https://doi.org/10.2308/acch.2008.22.1.69

Czerney, K., Schmidt, J. J., & Thompson, A. M. (2019). Do investors respond to explanatory language included in unqualified audit reports? *Contemporary Accounting Research*, *36*(1), 198–229. https://doi.org/10.1111/1911-3846.12425

Darendeli, A., Fiechter, P., Hitz, J.-M., & Lehmann, N. (2022). The role of corporate social responsibility (CSR) information in supply-chain contracting: Evidence from the expansion of CSR rating coverage. *Journal of Accounting and Economics*, *74*(2–3), Article 101525. https://doi.org/10.1016/j.jacceco.2022.101525

Davis, A. K., & Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, *29*(3), 804–837. https://doi.org/10.1111/j.1911-3846.2011.01125.x

Dennis, S. A., Griffin, J. B., & Zehms, K. M. (2019). The value relevance of managers' and auditors' disclosures about material measurement uncertainty. *The Accounting Review*, *94*(4), 215–243. https://doi.org/10.2308/accr-52272

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245. https://doi.org/10.1016/j.jacceco.2017.07.002

Elliott, W. B., Fanning, K., & Peecher, M. E. (2020). Do investors value higher financial reporting quality, and can expanded audit reports unlock this value? *The Accounting Review*, *95*(2), 141–165. https://doi.org/10.2308/accr-52508

European Commission. (2011, November 30). *Regulation of the European Parliament and of the Council on specific requirements regarding statutory audit of public-interest entities*. https://op.europa.eu/de/publication-detail/-/publication/6de7d1dc-7261-4eb8-ae57-5e9afbda7463/language-en

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Filip, A., Jeanjean, T., & Paugam, L. (2015). Using real activities to avoid goodwill impairment losses: Evidence and effect on future performance. *Journal of Business Finance & Accounting*, *42*(3–4), 515–554. https://doi.org/10.1111/jbfa.12107

Financial Reporting Council (FRC). (2013, February 1). *Revision to ISA (UK and Ireland) 700. Requiring the auditor's report to address risks of material misstatement, materiality and a summary of the audit scope*. https://www.frc.org.uk/getattachment/b567ab62-dfa5-4b61-a052-852c4bf51f0e/;.aspx

Financial Reporting Council (FRC). (2016). *Extended auditor's report: A further review of experience*. https://media.frc.org.uk/documents/Extended_auditors_reports_A_further_review_of_experience.pdf

Gimbar, C., Hansen, B., & Ozlanski, M. E. (2016). The effects of critical audit matter paragraphs and accounting standard precision on auditor liability. *The Accounting Review*, *91*(6), 1629–1646. https://doi.org/10.2308/accr-51382

Glaum, M., Landsman, W. R., & Wyrwa, S. (2018). Goodwill impairment: The effects of public enforcement and monitoring by institutional investors. *The Accounting Review*, *93*(6), 149–180. https://doi.org/10.2308/accr-52006

Gunn, J. L., Khurana, I. K., & Stein, S. E. (2018). Determinants and consequences of timely asset impairments during the financial crisis. *Journal of Business Finance & Accounting*, *45*(1–2), 3–39. https://doi.org/10.1111/jbfa.12287

Gutierrez, E., Minutti-Meza, M., Tatum, K. W., & Vulcheva, M. (2018). Consequences of adopting an expanded auditor's report in the United Kingdom. *Review of Accounting Studies*, *23*(4), 1543–1587. https://doi.org/10.1007/s11142-018-9464-0

Han, H., Tang, J. J., & Tang, Q. (2021). Goodwill impairment, securities analysts, and information transparency. *European Accounting Review*, *30*(4), 767–799. https://doi.org/10.1080/09638180.2020.1791725

Holzman, E. R., Miller, B. P., & Williams, B. M. (2021). The local spillover effect of corporate accounting misconduct: Evidence from city crime rates. *Contemporary Accounting Research*, *38*(3), 1542–1580. https://doi.org/10.1111/1911-3846.12659

Honkamäki, T., Mättö, M., & Teittinen, H. (2022). The homogeneity of BIG4 audit reports after the implementation of key audit matters in the context of fair value accounting. *International Journal of Auditing*, *26*(3), 354–370. https://doi.org/10.1111/ijau.12285

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, *64*(6), 2833–2855. https://doi.org/10.1287/mnsc.2017.2751

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

IAASB. (2013). *Reporting on audited financial statements: Proposed new and revised international standards on auditing*. International Federation of Accountants.

IAASB. (2015, January 1). *The new auditor's report: Greater transparency into the financial statement audit*. https://www.ifac.org/system/files/publications/files/Auditor-Reporting-Fact-Sheet.pdf

Ianniello, G., & Galloppo, G. (2015). Stock market reaction to auditor opinions—Italian evidence. *Managerial Auditing Journal*, *30*(6–7), 610–632. https://doi.org/10.1108/MAJ-06-2014-1045

IASB. (2022, October 6). *Who uses IFRS Accounting standards?* https://www.ifrs.org/use-around-the-world/use-of-ifrs-standards-by-jurisdiction/

Jahan, N., & Karim, M. S. (2024). Does goodwill-related critical audit matters disclosure influence firms' financial reporting decisions? Evidence from goodwill impairment. *Auditing: A Journal of Practice & Theory*, *43*(2), 159–187. https://doi.org/10.2308/AJPT-2022-129

Kabir, H., & Rahman, A. (2016). The role of corporate governance in accounting discretion under IFRS: Goodwill impairment in Australia. *Journal of Contemporary Accounting & Economics*, *12*(3), 290–308. https://doi.org/10.1016/j.jcae.2016.10.001

Kachelmeier, S. J., Rimkus, D., Schmidt, J. J., & Valentine, K. (2020). The forewarning effect of critical audit matter disclosures involving measurement uncertainty. *Contemporary Accounting Research*, *37*(4), 2186–2212. https://doi.org/10.1111/1911-3846.12583

Kim, O., & Verrecchia, R. E. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research*, *29*(2), 302–321. https://doi.org/10.2307/2491051

Klevak, J., Livnat, J., Pei, D. S., & Suslava, K. (2023). Critical audit matters: Possible market misinterpretation. *Auditing: A Journal of Practice & Theory*, *42*(3), 45–70. https://doi.org/10.2308/AJPT-2020-113

Köhler, A., Ratzinger-Sakel, N., & Theis, J. (2020). The effects of key audit matters on the auditor's report's communicative value: Experimental evidence from investment professionals and non-professional investors. *Accounting in Europe*, *17*(2), 105–128. https://doi.org/10.1080/17449480.2020.1726420

Lennox, C. S., Schmidt, J. J., & Thompson, A. M. (2023). Why are expanded audit reports not informative to investors? Evidence from the United Kingdom. *Review of Accounting Studies*, *18*, 497–532. https://doi.org/10.1007/s11142-021-09650-4

Li, H., Hay, D., & Lau, D. (2019). Assessing the impact of the new auditor's report. *Pacific Accounting Review*, *31*(1), 110–132. https://doi.org/10.1108/PAR-02-2018-0011

Li, V., & Luo, Y. (2023). Costs and benefits of auditors' disclosure of critical audit matters: Initial evidence from the United States. *Advances in Accounting*, *60*, Article 100641. https://doi.org/10.1016/j.adiac.2022.100641

Li, Z., Shroff, P. K., Venkataraman, R., & Zhang, I. X. (2011). Causes and consequences of goodwill impairment losses. *Review of Accounting Studies*, *16*(4), 745–778. https://doi.org/10.1007/s11142-011-9167-2

Liao, L., Minutti-Meza, M., Zhang, Y., & Zou, Y. (2022). *Consequences of the adoption of the expanded auditor's report: Evidence from Hong Kong*. Working paper. https://doi.org/10.2139/ssrn.3392449

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Minutti-Meza, M. (2021). The art of conversation: The expanded audit report. *Accounting and Business Research*, *51*(5), 548–581. https://doi.org/10.1080/00014788.2021.1932264

Mock, T. J., Bédard, J., Coram, P. J., Davis, S. M., Espahbodi, R., & Warne, R. C. (2013). The audit reporting model: Current research synthesis and implications. *Auditing: A Journal of Practice & Theory*, *32*(Supplement 1), 323–351. https://doi.org/10.2308/ajpt-50294

Moroney, R., Phang, S.-Y., & Xiao, X. (2021). When do investors value key audit matters? *European Accounting Review*, *30*(1), 63–82. https://doi.org/10.1080/09638180.2020.1733040

Myers, L. A., Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2018). Measuring the market response to going concern modifications: The importance of disclosure timing. *Review of Accounting Studies*, *23*(4), 1512–1542. https://doi.org/10.1007/s11142-018-9459-x

Ogneva, M., & Subramanyam, K. R. (2007). Does the stock market underreact to going concern opinions? Evidence from the US and Australia. *Journal of Accounting and Economics*, *43*(2–3), 439–452. https://doi.org/10.1016/j.jacceco.2006.12.004

PCAOB. (2017, June 1). *The auditor's report on an audit of financial statements when the auditor expresses an unqualified opinion and related amendments to PCAOB standards*. https://pcaobus.org/Rulemaking/Docket034/2017-001-auditors-report-final-rule.pdf

Porumb, V., Zengin-Karaibrahimoglu, Y., Lobo, G. J., Hooghiemstra, R., & Waard, D. (2021). Expanded auditor's report disclosures and loan contracting. *Contemporary Accounting Research*, *38*(4), 3214–3253. https://doi.org/10.1111/1911-3846.12697

Ramanna, K., & Watts, R. L. (2012). Evidence on the use of unverifiable estimates in required goodwill impairment. *Review of Accounting Studies*, *17*(4), 749–780. https://doi.org/10.1007/s11142-012-9188-5

Rapley, E. T., Robertson, J. C., & Smith, J. L. (2021). The effects of disclosing critical audit matters and auditor tenure on nonprofessional investors' judgments. *Journal of Accounting & Public Policy*, *40*(5), Article 106847. https://doi.org/10.1016/j.jaccpubpol.2021.106847

Rousseau, L. M., & Zehms, K. M. (2024). It's a matter of style: The role of audit firms and audit partners in key audit matter reporting. *Contemporary Accounting Research*, *41*(1), 529–561. https://doi.org/10.1111/1911-3846.12902

Ruhnke, K., & Schmidt, M. (2014). The audit expectation gap: Existence, causes, and the impact of changes. *Accounting and Business Research*, *44*(5), 572–601. https://doi.org/10.1080/00014788.2014.929519

Scancell Holdings plc. (2018). *Report and consolidated financial statements for the year ended 30 April 2018*. https://scancell.co.uk/wp-content/uploads/2024/08/scancell-holdings-accounts-30042018.pdf

Seebeck, A., & Kaya, D. (2023). The power of words: An empirical analysis of the communicative value of extended auditor reports. *European Accounting Review*, *32*(5), 1185–1215. https://doi.org/10.1080/09638180.2021.2021097

Siano, F. (2022). *Contextualized news in corporate disclosures: A neural language approach*. https://hdl.handle.net/10125/103979

Siano, F., & Wysocki, P. (2021). Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small(er) datasets. *Accounting Horizons*, *35*(3), 217–244. https://doi.org/10.2308/HORIZONS-19-161

Siano, F., & Wysocki, P. D. (2018). *The primacy of numbers in financial and accounting disclosures: Implications for textual analysis research*. Working paper. https://doi.org/10.2139/ssrn.3223757

Sikka, P. (2009). Financial crisis and the silence of the auditors. *Accounting, Organizations and Society*, *34*(6–7), 868–873. https://doi.org/10.1016/j.aos.2009.01.004

Sirois, L.-P., Bédard, J., & Bera, P. (2018). The informational value of key audit matters in the auditor's report: Evidence from an eye-tracking study. *Accounting Horizons*, *32*(2), 141–162. https://doi.org/10.2308/acch-52047

Smith, K. W. (2023). Tell me more: A content analysis of expanded auditor reporting in the United Kingdom. *Accounting, Organizations and Society*, *108*, Article 101456. https://doi.org/10.1016/j.aos.2023.101456

Soltani, B. (2000). Some empirical evidence to support the relationship between audit reports and stock prices—The French case. *International Journal of Auditing*, *4*(3), 269–291. https://doi.org/10.1111/1099-1123.00317

Song, S. (2021). The informational value of segment data disaggregated by underlying industry: Evidence from the textual features of business descriptions. *The Accounting Review*, *96*(6), 361–396. https://doi.org/10.2308/TAR-2017-0572

Tecan Group Ltd. (2021). *Tecan annual report 2020: Connecting—People and partners empowered to fight COVID-19*. https://www.tecan.com/annual-reports

Vinci SA. (2021). *Forging a sustainable world: 2020 universal registration document*. https://www.vinci.com/en/finance/investors/regulatory-information/reports-and-presentations?f%5B0%5D=annee%3A2020&page=1

Watchstone Group plc. (2018). *Annual report and financial statements 2017*. https://www.annualreports.com/HostedData/AnnualReportArchive/w/ASX_WTG_2017.pdf

Yang, Y., Uy, M. C. S., & Huang, A. (2020). *FinBERT: A pretrained language model for financial communications*. https://doi.org/10.48550/ARXIV.2006.08097

Yip, P. S. L., & Tsang, E. W. K. (2007). Interpreting dummy variables and their interaction effects in strategy research. *Strategic Organization*, *5*(1), 13–30. https://doi.org/10.1177/1476127006073512

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Küster, S., Steindl, T., & Göttsche, M. (2025). The informational content of key audit matters: Evidence from using artificial intelligence in textual analysis. *Contemporary Accounting Research*, *42*(4), 2392–2423. https://doi.org/10.1111/1911-3846.13070

APPENDIX

Variable	Description	Source
Experimental variables		
KAM_IMP_PROB	Predicted impairment probability for firm *i* in year $t + 1$ based on the FinBERT classification algorithm with the full KAM of firm *i* in year *t* as the input text sequence	Own calculation
DESC_IMP_PROB	Predicted impairment probability for firm *i* in year $t + 1$ based on the FinBERT classification algorithm with the description section of the KAM of firm *i* in year *t* as the input text sequence	Own calculation
RESP_IMP_PROB	Predicted impairment probability for firm *i* in year $t + 1$ based on the FinBERT classification algorithm with the response section of the KAM of firm *i* in year *t* as the input text sequence	Own calculation
KAM_IMP_SIZE	Predicted impairment amount for firm *i* in year $t + 1$ based on the FinBERT continuous prediction algorithm with the full KAM of firm *i* in year *t* as the input text sequence	Own calculation
DESC_IMP_SIZE	Predicted impairment amount for firm *i* in year $t + 1$ based on the FinBERT continuous prediction algorithm with the description section of the KAM of firm *i* in year *t* as the input text sequence	Own calculation
RESP_IMP_SIZE	Predicted impairment amount for firm *i* in year $t + 1$ based on the FinBERT continuous prediction algorithm with the response section of the KAM of firm *i* in year *t* as the input text sequence	Own calculation
FUNDA_IMP_PROB	Impairment probability of firm *i* for $t + 1$ modeled out-of-sample using all control variables and a Gradient Boosting machine learning algorithm	Own calculation
DISC_IMP_PROB	Impairment probability of firm *i* for $t + 1$ modeled out-of-sample with a FinBERT prediction model using goodwill-related disclosures within the firm's annual report as the input text sequence	Own calculation
Dependent variables		
CAR	Cumulative abnormal returns in the 3-day period surrounding the audit opinion date calculated as the security return minus the return of the overall market	LSEG
GC_OPINION	Indicator variable that equals one if firm *i* receives a going concern opinion in year $t + 1$, and zero otherwise	Audit Analytics
IMPAIR	Indicator variable that equals one if firm *i* recognizes an impairment of goodwill in year $t + 1$, and zero otherwise	LSEG
IMPAIR_SIZE	Absolute amount of goodwill impairment scaled by total assets	LSEG
RESTR	Indicator variable that equals one if firm *i* recognizes restructuring expenses in year $t + 1$, and zero otherwise	LSEG
Control variables		
BATH	Indicator variable that equals one if firm i's operating income in year *t* is negative and the decline in operating income is below the median among firms with a negative change in operating income in that year, and zero otherwise	LSEG
CEO_COMP	Indicator variable that equals one if the compensation of the CEO of firm *i* is linked to total shareholder return in year *t*, and zero otherwise	LSEG

APPENDIX (Continued)

Variable	Description	Source
FOLLOWING	Natural logarithm of the total number of analysts following the firm	LSEG
GW_RATIO	Gross goodwill divided by total assets before goodwill impairment of firm *i* in year *t*	LSEG
IMPAIR_HIST	Indicator variable that equals one if firm *i* recognized an impairment of goodwill in year *t*, and zero otherwise	LSEG
LARGE	Indicator variable that equals one if a firm's market capitalization is higher than the sample median, and zero otherwise	Own calculation
LEVERAGE	Total debt divided by total assets of firm *i* in year *t*	LSEG
MTB	Market capitalization divided by the total book value of firm *i* in year *t*	LSEG
NUM_KAM	Number of KAMs reported in the audit report of firm *i* in year *t*	Audit Analytics
RETURN	Total stock return for firm *i* in year *t* adjusted by the MSCI Global total stock return in year *t*	LSEG
ROA	ROA of firm *i* in year *t* adjusted by the industry-year median of ROA in year *t*	LSEG
SEGMENTS	Natural logarithm of 1+ number of reported business segments of firm *i* in year *t*	LSEG
SIZE	Natural logarithm of market capitalization of firm *i* in year *t*	LSEG
SMALL	Indicator variable that equals one if a firms' market capitalization is lower than or equal to the sample median, and zero otherwise	Own calculation
SMOOTH	Indicator variable that equals one if firm *i*'s operating income in year *t* is positive and the increase in operating income is above the median among firms with a positive change in operating income in that year, and zero otherwise	LSEG
SPECIAL_ITEMS	Ratio of nonrecurring income/expense divided by total revenues	LSEG
SURPRISE	Difference between the actual earnings per share and the last mean forecast of the period, expressed as a percentage	LSEG