# Empowering Intermediaries: Machine Learning for Sustainable Business Model Innovation in SMEs

**Dieke Marlies Koers-Stuiver**

*Saxion University of Applied Sciences*
d.m.koers@saxion.nl

**Mark de Lat**
*Saxion University of Applied Sciences, Nyenrode Business University*
m.b.j.delat@saxion.nl

**Timber Haaker**
*Saxion University of Applied Sciences*
t.i.haaker@saxion.nl

## ABSTRACT

Small and medium-sized enterprises (SMEs) are crucial for achieving sustainable transitions; however, they often find business model innovation for sustainability (BMfS) complex. Intermediaries, such as organizational advisors, play a crucial role in facilitating this transition, yet they face challenges arising from fragmented sustainability concepts, diverse business model tools (BMT), contextual needs, and an absence of clear conceptual terminology, such as a lack of shared language. This study, therefore, explores how machine learning (ML), specifically natural language processing (NLP), can enable intermediaries to assist SMEs more effectively in their sustainable transitions. Through a mixed-methods longitudinal case study involving two Dutch intermediary organizations, we investigate how ML-based insights can enhance contextual knowledge and foster shared language creation. We conduct a mapping review of over 900 documents, utilizing topic modeling techniques alongside qualitative group interviews to examine how ML-driven tools can consolidate fragmented sustainability concepts and customize BMTs for sustainability for SME contexts. Preliminary findings suggest the viability of a hybrid advisory model that integrates theory-informed frameworks with ML-enhanced contextual analysis, providing intermediaries with improved capabilities to identify sustainability challenges, engage stakeholders, and accelerate BMIfS

## KEYWORDS

Machine Learning, Natural Language Processing, Business Model Innovation, Sustainability, Intermediaries.

# INTRODUCTION

SMEs face challenges in transitioning their business models (BMs) towards sustainability, as this often involves complex change processes and business model innovation (Moursellas et al. 2023; Pinkse et al. 2023). This difficulty primarily arises from the wicked nature and context-dependent challenges that require innovative approaches to value creation. Collaboration with various stakeholders is often limited by constrained resources and a lack of expertise transformations (Pinske, 2023). A major challenge for small and medium-sized enterprises (SMEs) is determining the right demarcation of the various sustainability objectives. Sustainability is a multifaceted concept explored by numerous scholars and encompasses various frameworks, including, but not limited to, circular economy principles, Environmental, Social, and Governance (ESG), multiple value creation, Corporate Social Responsibility (CSR), blended or hybrid value creation (Emerson 2003; Geissdoerfer et al. 2020; Joyce and Paquin 2016; Midgley and Arya 2022). This concept's unclear and fragmented interpretation hinders sustainable transitions, resulting in further fragmentation of the sustainability dimensions. Additionally, numerous domains offer tailored templates, tools, and strategies aimed at particular sustainability aspects (Koers-Stuiver et al. 2023). These conceptual silo's contribute to a less rapid sustainable transition. While SMEs are vital for reaching sustainability objectives, there is a significant lack of research focused on them (Carlsson and Nevzorova 2025; Khoja et al. 2022; Salvador et al. 2023).

This is why this research focuses on the role of intermediaries due to their essential support for SME clients (Quartey and Oguntoye 2020). Intermediaries are vital in aiding SMEs as they navigate the complexities of adopting BMfS through various tools and methodologies (Salvador 2023). Intermediaries are essential in promoting and accelerating sustainable transitions in SMEs(Kanda et al. 2018; Mignon and Kanda 2018). They play a vital role in the development, design, or implementation of new processes, products, or practices(Kundurpi et al. 2021). Intermediaries connect actors, activities, and resources, aligning network visions with established regimes. They foster collaboration among niche technologies and markets, creating momentum for socio-technical system changes and challenging unsustainable practices setups (Kivimaa et al. 2019a; Lobo et al. 2025; Murto et al. 2020). They are crucial in encouraging sustainable transitions by building ecosystems where they articulate, negotiate, and align varied interests across different niches and regimes (Zhang,2020). During change initiatives, intermediaries frequently utilize BMTs to navigate complexity and support the design, development, and implementation of BMfS. Navigating the challenges posed by a fast-changing market, especially regarding BMT and contextual sustainability metrics, poses challenges for intermediaries (Breuer et al., 2018; Kivimaa et al., 2020). Context includes information about specific situations, people, and objects essential for understanding it (Messmer et al. 2024). Furthermore, this evolution and the need for interdisciplinary expertise influence intermediaries' ability to manage these shifts, emphasizing the importance of language (Sancak, 2023).

Therefore, this research sets out to answer the following: How can insights from ML, including NLP, enhance intermediaries' ability to support business model innovation for sustainability (BMIfS) in SMEs through shared language and context understanding? To answer this question, we set out to map and synthesize the sustainability dimensions and influence of contextual and shared language creation on BMTfS, using ML to enhance the relevance of sustainability goals. It aims to provide intermediaries with tailored tools and insights via ML for faster BMIfS. In doing so, we aim to enhance advisory practices and support SMEs in their sustainable transitions.

## BACKGROUND

Digital technologies such as AI, ML and NLP enhance sustainable value creation and can stimulate BMI (Aagaard and Tucci 2024; Divyashree et al. 2024; Moghrabi et al. 2023). BMI takes place when an organization's key components and business logic are deliberately modified, influencing how it generates, captures, and provides value (Amit and Zott 2012; Foss and Saebi 2017). BM overviews often include the value proposition, resources, key activities, cost and revenue model, value chain, and target market((Aagaard 2024; Massa et al. 2017; Osterwalder and Pigneur 2010; Teece 2010). BMI can result in creating completely new BMs, diversifying into different ones, acquiring new models, or shifting from one model to another (Geissdoerfer et al. 2018).

BMI for sustainability (BMfS) facilitates change in balancing and creating multiple values (Evans et al. 2017). This entails simultaneously seeking long-term social, ecological, and economic value while taking into account all relevant stakeholders (Aagaard 2024; Koers-Stuiver et al. 2024; Ritala et al. 2021). This complex and multidimensional process involves challenging applications that require diverse approaches and contextual insights across disciplines, as different perspectives influence the understanding of sustainable BMT (Haynes and Alemna 2022; Velter et al. 2022). Contextual insights refer to information related to specific situations, individuals, objects, and entities necessary for understanding them (Messmer et al. 2024). This leads to a demand for interdisciplinary practical tools and strategies to effectively implement Business Models for Sustainability (BMfS). As a result, SMEs frequently seek guidance from intermediaries (Kanda et al. 2018; Köhler et al. 2019)

Intermediaries are essential in promoting and accelerating sustainable transitions in SMEs (Kanda et al. 2018; Mignon and Kanda 2018). Intermediaries play a vital role in the development, design, or implementation of new processes, products, or practices (Kundurpi et al. 2021). They can include various entities such as government agencies, business networks, academic institutions, expert civil society organizations, and consultancy firms (Lord et al. 2025; Passaro et al. 2023), of which the latter is the focus of this research. As such, intermediaries play a crucial role in positively shaping sustainability transitions(Kivimaa et al. 2019b). Sustainable transitions refer to fundamental shifts in society's functions, intending to replace existing practices with new sustainable systems (Köhler et al. 2019; Lord et al. 2025). Intermediaries drive systemic change by mobilizing resources, offering strategic guidance on technology adoption, and promoting collaboration among stakeholders with complementary strengths (Mignon and Kanda 2018). For SMEs, intermediaries are vital not just at the beginning but throughout the transition, as they engage, manage cooperation costs, and ensure that long-term sustainability goals take precedence over immediate aims (Iturrioz et al. 2015). Furthermore, they facilitate the development and commercialization of ecological innovations, nurture niche market growth, and champion favorable policy and regulatory conditions (Mignon and Kanda 2018). Their support strategies are highly contextual, combining both general and customized approaches to help organizations navigate the complexities of sustainable innovation (Iturrioz et al. 2015; van Lente et al. 2020).

During change initiatives, intermediaries use business BMTs to support the design and implementation of BMfS. BMT are systematic approaches that enable organizations to create, assess, and refine their BMs. The landscape of BMT is complex, featuring diverse tools and dimensions for value creation and BMI (Koers-Stuiver et al. 2024). Balancing the fast-evolving market demands, especially regarding business tools and sustainability metrics, while effectively allocating resources to help clients achieve their sustainability goals, is challenging for intermediaries (Breuer et al. 2018; Kivimaa et al. 2020). Additionally, the rapid pace of change and the need for interdisciplinary

knowledge can affect intermediaries' ability to navigate these transitions, where language is crucial(Sancak 2023).

To thrive, intermediaries must effectively construct a shared language and understand the specific SME's context. This highlights the importance of straightforward, user-friendly BMTs that can be customized to meet contextual demands and stakeholder needs(Bhatnagar et al. 2022; Kanda et al. 2018). Language diversity further exacerbates the complexity, impacting both individuals and groups and possibly hindering the interpretability of tools and value-creation concepts (Lecomte et al. 2023) Language can function as a system of meaning essential for understanding organizational, social, and global contexts (Sanden 2016). Sanden (2016) further asserts that language transcends mere communication; it embodies both identity and power, influencing organizational structures and power dynamics. Neglecting the impact of linguistic diversity can lead to inefficiencies, hinder collaboration, and reduce productivity. Therefore, understanding how the development of a shared language and the impact of context affect intermediaries and their clients in utilizing BMT is essential.

Shared language construction encompasses various elements that affect the usage, creation, and interpretation of language (Bera et al. 2014; Holmes and Wilson 2022; Spivey et al. 2012; Wen and Taylor 2021). For language construction to be effective, it must offer practical and relevant content that resonates with individual employees' and the organization's existing knowledge and goals(Baker and Welter 2020). Tools are most effective when users grasp the language and terminology involved (Szopinski et al. 2022). The comprehensibility of BMT language is shaped by personal factors such as cognitive abilities, language processing skills, and prior knowledge of the user (Baron 2006). Furthermore, experience plays a crucial role in language construction, which includes industry exposure, an understanding of the ecosystem, familiarity with the supply chain, the type of organization, the composition of the management team, company size, and corporate culture. The context and the way language is created greatly affect user interactions with SBM tools; however, this area remains underexplored. AI, ML, and NLP can help close this gap.

AI allows businesses to create new value, increase efficiency, and gain competitive advantage (Aagaard & Tucci, 2024). Moreover, it helps to recognize hidden patterns and contextual meanings in sustainability transformation-related information (Gamlath et al. 2023; Telukdarie et al. 2024). The widespread use of digital platforms has led to a surge in textual data, including emails, reports, social media exchanges, and customer feedback (Milei et al. 2025; Singh 2023). Similarly, the number of published research papers in various domains has grown exponentially. This increase in text-based organizational interactions and scholarly publications has outpaced manual processing capabilities, leading to algorithms such as ML and NLP to extract valuable insights (Medeiros et al. 2020). ML, a branch of AI, aims to develop algorithms that enable computers to learn from data and execute tasks like classification, prediction, clustering, and knowledge extraction without direct programming (Grimmer et al. 2022). Additionally, ML applications are increasingly integrating contextual awareness into their algorithms, resulting in insights about specific situations, individuals, objects, and entities crucial for understanding and interpretation. (Messmer et al., 2024). Meanwhile, NLP is a field of AI that enables computers to comprehend, interpret, and produce human language. It serves as a bridge between human communication and computer understanding, allowing machines to engage with humans more naturally and intuitively (Chauhan and Shah 2021; Kang et al. 2020). The growth of ML and NLP methods offers new possibilities for analyzing large volumes of text from numerous online business sources and important research in multiple disciplines, facilitating the discovery of hidden meanings in language and its contextual influences (Montejo-Ráez and Jiménez-Zafra 2022).

Intermediaries are increasingly utilizing data-driven methods, acknowledging data's crucial role in decision-making and organizational change (Adaga et al., 2024; Reddy, 2025). When supported by digital technologies, data generates value for stakeholders and propels BMIfS forward (Stahl et al., 2023). The application of AI and ML greatly improves operational efficiency, offers predictive insights, and aids client decision-making (Osman et al., 2025). NLP specializes in analyzing unstructured data such as social media to uncover customer sentiments, which facilitates personalized communication (Reddy, 2025). AI also plays a role in change management by assessing performance, processes, and customer feedback to pinpoint areas needing transformation (Mohan, 2024). Intermediaries face challenges in implementing ML, including data integrity, ethical concerns, and integration difficulties (Osman et al., 2025). They connect technical solutions with organizational contexts, facilitating necessary behavioral and structural changes (Reddy, 2025). Moreover, ML supports agile strategy development by detecting market trends and risks (Osman et al., 2025), thus enhancing intermediation processes (Di Silva, 2018). The conceptual model shown in Figure 1 demonstrates how intermediaries utilize insights from ML and NLP to integrate contextual knowledge and shared language creation. This integration allows intermediaries to customize BMT for particular sustainability aspects, enabling BMIfS in small and SMEs.

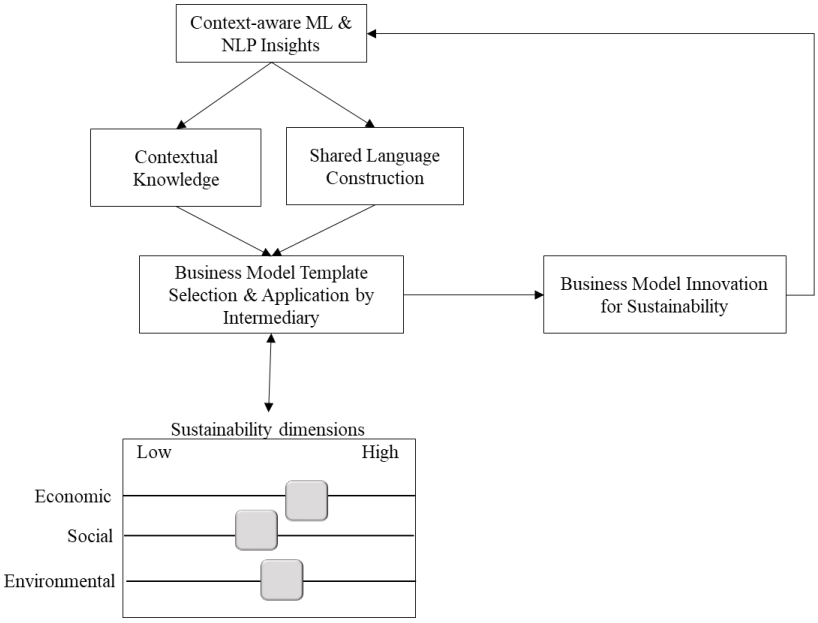


*Figure 1 Conceptual Model*

## METHOD

Our study uses a mixed-methods approach to examine the influence of language construction and the context on BMTfS in the advisory process of two Dutch intermediaries. Our research incorporates participatory principles, ML methodologies, and content analysis to enhance our understanding of this process (Burns et al. 2021; Creswell 2003; Tashakkori and Teddlie 1998). We adopt a

longitudinal case study research design (Gerring 2007) to investigate the advisory process of two Dutch intermediaries and their clients.

## Case Organization

Both organizations function as intermediaries, assisting SMEs in incorporating sustainability principles throughout their operations. These organizations are located in the eastern part of the Netherlands. With considerable expertise in SME collaborations, these organizations share a focus on generating positive social and environmental value. While Eshuis Accounting en Adviseurs emphasizes CSRD, and de Betekenisfabriek concentrates on Identity marketing, both are key changemakers in the complex shift towards sustainability. Their specialized knowledge allows them to accurately identify organizational challenges and needs, facilitating the development of tailored solutions. This project started in November 2024.

## Group interviews

This study employs a year-long iterative data collection process involving semi-structured group interviews(Wilkinson 1998). Please see Table 1 for the scheduled date of the group interview with the intermediaries, along with its objectives, duration, and key concepts.

*Table 1. Partner meeting overview*

| Date | Focus | Duration | Type | Available documentation |
|---|---|---|---|---|
| 13-12-2024 | Kick-Off | 90 minutes | Physical | Preparation materials, minutes |
| 11-03-2025 | Refinement research scope | 60 minutes | Online | Preparation materials, minutes, transcript |
| 04-04-2025 | Refinement RM | 60 minutes | Online | Preparation materials, minutes, transcript |
| End of April | Data analysis pipeline | Tbd | Tbd | n.a |

## Mapping review using NLP methods

We employ machine learning and natural language processing to conduct a mapping review of the literature pertinent to our research question and objectives (Asmussen and Møller 2019; Duriau et al. 2007; Munn et al. 2018; Tricco et al. 2018). This review centers on gathering and analyzing literature regarding the development of shared language, context, sustainable business model innovation, intermediaries, and related synonyms for these constructs.

Our research has encompassed the collection of over 900 documents, containing utilize ML and NLP to review literature. This process involves gathering and analyzing over 900 scientific articles and books, as shown in Table 2. We used purposive sampling to identify keywords linked to our queries and examined related documents and citations. The databases for our investigation included Scopus, Web of Science, Google Scholar and SpringerLink, along with undermine.ai and typeset.io for additional sources. Each document was assessed based on its title, keywords, and abstract, with selection criteria focusing on the concepts, methodologies, research gaps, and future directions. Selected documents were converted to .txt for efficient processing and organized with titles containing Year, First Author, and Title, using an algorithm to detect duplicates.

To find the contextual meaning of the key terms, we created a comprehensive folder of 933 relevant documents, categorized into subfolders by specific keywords. This led to overlaps, as documents may belong to multiple thematic categories like ESG, NLP, SUS, and BM. We are currently analyzing the subfolders.

*Table 2. Document folder based on themes*

| Folder | Keywords | N Docs |
|---|---|---|
| ESG (Environmental, Social, Governance) | ESG, Environmental, Social, Governance, CSR, Corporate Social Responsibility, CSRD, Corporate Social Responsibility Directive, Blended Value, People, Planet, Profit, Triple Bottom Line, Triple Layer | 90 |
| BM (Business Model) | Business Model, Business Models | 289 |
| BMT (Business Model Tool) | Business Models Template, Business Model Canvas, Business Model Method, Business Model Model, Triple Layer Canvas, Business Model Tool, Business Model Taxonomy, Business Model Framework, Business Model Method, Business Model Taxonomy | 61 |
| SUS (Sustainability) | Sustainability, Green, Ecological, Environmental, Circular | 286 |
| INT (Intermediary AND sustainability OR ML) | Intermediaries, Advisors, Consultant(s), Agent, Management Consult*, | 54 |
| LANG (Language) | Language, Narrative, Language Creation, Shared Language, Language Construction, Conversation, Semantic | 61 |
| ML (Machine learning | Machine learning, Artificial Intelligence, NLP, Natural Language Processing, Topic Modeling, Text Mining, Text Analytics | 217 |

*Software used*

At the start of our research project, we first used RStudio (4.13). However, in early April 2025, we switched to Jupyter Lab (Python 3.10) because the processing time in RStudio became too cumbersome for our analysis. This analyses are performed in Python, utilizing libraries such as nltk, pandas, scikit-learn, matplotlib, and bertopic. The environment was managed with conda, ensuring that key dependencies were installed as required

*Topic modeling*

Latent Dirichlet Allocation (LDA) was utilized through Gensim's LdaModel to identify hidden thematic structures within the corpus. The sparse matrix was transformed into Gensim's corpus format via Sparse2Corpus for compatibility. Preliminary exploratory analysis guided the setting of the number of topics and other hyperparameters. The outcomes of topic modeling were visualized

using pyLDAvis, which provides an interactive interface to explore topic distributions and key terms associated with each topic.

Furthermore, we employed the BERTopic framework to analyze the results (Grootendorst, 2022). BERTopic is a topic modeling approach that utilizes pre-trained transformer-based embeddings along with class-based TF-IDF (c-TF-IDF) to produce interpretable topics. In contrast to traditional models like LDA, BERTopic enables a more sophisticated semantic analysis of documents by using contextual embeddings (all-MiniLM-L6-v2 from SentenceTransformers) to group semantically similar text segments. Following the initial modeling, we optionally reduced the number of topics to enhance coherence and interpretability. Topic labels were created from the highest-ranked terms for each topic and were consistently applied across all visualizations. We examined temporal trends by matching each document chunk with its original publication year and utilized BERTopic's topicsovertime method. Additionally, visualizations such as UMAP projections and similarity heatmaps were produced to investigate topic relationships and their distribution.

*Data preparation en preprocessing*

Our study employed NLP techniques to examine a set of texts aligned with our research goals. The workflow included loading texts, cleaning data, lemmatizing, vectorizing, and performing topic modeling through BERTopic (And LDAVIS). We gathered documents from a designated folder filled with .txt files. For each document, we recorded metadata such as document IDs and estimated publication years, derived from filenames using regular expressions. Text preprocessing began with ASCII encoding to remove non-standard characters, followed by tokenization using nltk.wordtokenize, and the elimination of punctuation and stopwords.

We created a custom stopword list by merging standard English stopwords sourced from NLTK. Words underwent lemmatization with WordNetLemmatizer, applying POS tags from NLTK's postag to align them with WordNet categories. The processed texts were transformed into document-term matrices using CountVectorizer from scikit-learn. Topic modeling was then performed with the BERTopic framework, which combines transformer-based embeddings with clustering algorithms like HDBSCAN to yield interpretable topic representations. The resulting topics were visualized and saved for further analysis.

# PRELIMINARY FINDINGS

## Group Interviews

The content analysis of the three meeting transcripts identified several interconnected thematic areas that support the growing partnership among collaborators developing AI tools for sustainable business transformation.

Framing has become a crucial conceptual term in project discussions. Participants from De Betekenisfabriek emphasized the significance of identity frames, issue frames, and role-expectation frames as frameworks that influence organizational self-perception and stakeholder interactions. Although framing was already part of their approach, it was suggested that its role as a formal advisory tool, especially regarding CSRD implementation, could be further examined with partners such as Eshuis. This led to a discussion of the multiple application areas for this concept, rendering it less appropriate in highly structured settings where changes follow more standardized criteria.

A thematic difference has emerged between De Betekenisfabriek's framing-focused methods and Eshuis's compliance-driven efforts regarding the Corporate Sustainability Reporting Directive (CSRD). While both organizations operate in language-rich environments, their interpretive backgrounds and intervention strategies vary significantly. These discrepancies emphasize the need for a unified vocabulary and an aligned analytical framework to promote interoperability among different organizational domains.

In the meetings, the partners acknowledged a difference between language treated as content, which can be analyzed, and language models viewed as methods, including NLP and topic modeling. This distinction was vital for synchronizing the project's epistemological framework with its methodological objectives. Conversations emphasized the importance of clearly distinguishing language as a phenomenon influencing advisory processes from language models, which serve as technical instruments for data analysis.

Additionally, in meetings, the term tool was employed in various contexts, referencing both conceptual aids and tangible ML-driven applications. It also indicated interventions intended to modify business model components, whether communicated directly to the client or applied indirectly during process interpretation by the intermediaries. A common understanding emerged around the notion of BMT as a structured and flexible framework incorporating linguistic, organizational, and technological elements to facilitate sustainability transitions in SME organizations.

The participants focused on creating a data pipeline for this project in the session. They also suggested using ML techniques such as topic modeling, hierarchical clustering, and semantic similarity mapping to identify underlying frames, sentiment structures, and behavioral typologies from textual data such as daybooks, INK sessions, and websites. Early demonstrations showed that annotated corpora could aid in developing classification systems and predictive models informed by organizational language use.

It was acknowledged that effective ML methods necessitate both exploratory (unsupervised) and deductive (supervised) approaches. This requires the development of annotated datasets to train machine learning models on pertinent frames and dimensions of sustainability. Prototypes employing factor-based clustering and dimensional mapping were presented to exemplify how such models could facilitate advisory practices. The capacity of machine learning to generate domain-specific typologies and support reframing processes was confirmed, contingent upon the availability of structured and high-quality input data.

These preliminary findings suggest that combining narrative framing with data-driven modeling effectively promotes BMIfS in the advisory process. They also highlight the necessity of using clear language and shared language construction as interpretation and jargon rely significantly on context. Visualizations have emerged as a useful resource for establishing this common language framework in these meetings. The intersection of disciplines provides both the empirical and theoretical foundation for the tool being created in this project.

**Mapping Review**

In the following section, we present a part of our analysis, concentrating on two subsets (ESG and LANG) to exemplify the analytical potential.

*BERTopic*

This section presents a subset of our findings. We are currently working on refining the parameters; consequently, these visualizations may be subject to updates.

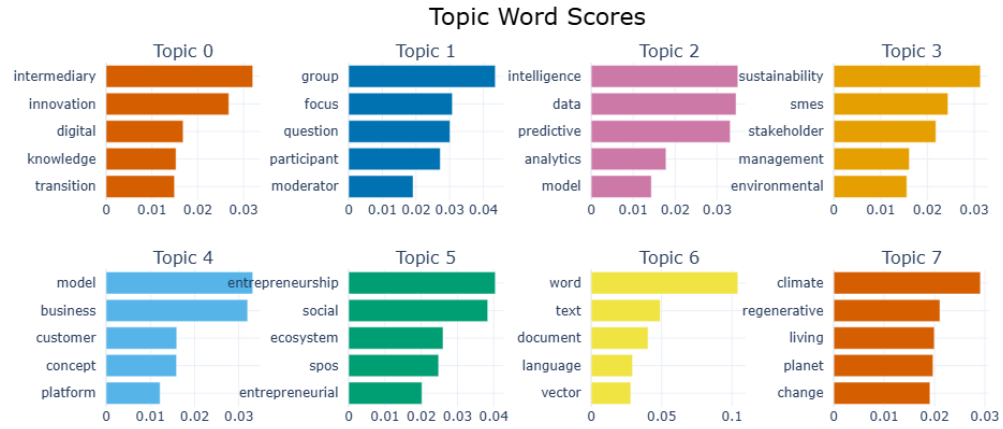


*Figure 2 INT Sample BERTopic Barchart*

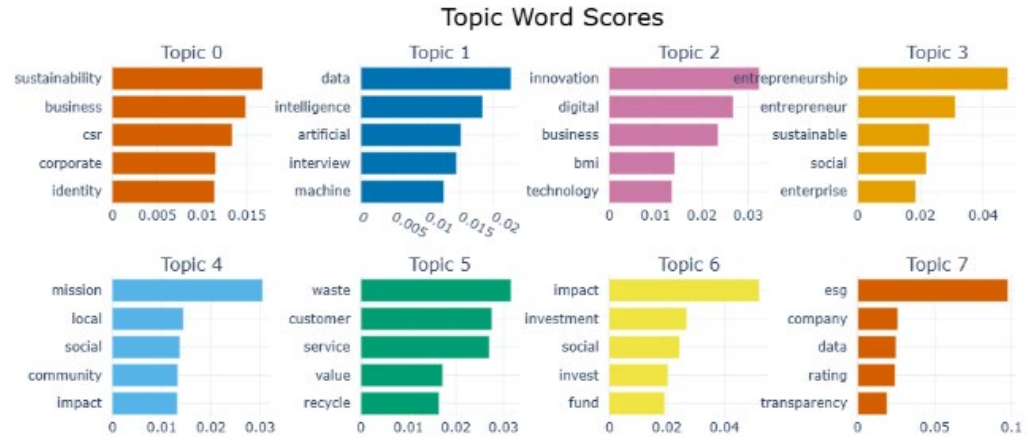


*Figure 3 ESG Sample BERTopic Barchart*

The INT sample captures broader conceptual themes of intermediaries, innovation, group collaboration, and sustainability transitions. The topics imply a strategic perspective, with terms like knowledge, transition, ecosystem, and entrepreneurship highlighting systems-level thinking. Topic 1 seems to focus on the role intermediaries have in change processes. Topics 2 and 6, which focus on data and intelligence, words, text, and language, highlight the literature on ML and NLP in sustainability transitions concerning intermediaries. The diversity among the topics emphasizes the exploratory and multi-stakeholder characteristics of intermediary discourse.

In comparison, the ESG sample appears more focused and applicable. It emphasizes particular sustainability areas such as CSR, identity, and investment effects and circular economy practices such as waste management, recycling, and services. Terms such as rating, transparency, company, and data

underscore performance-driven, compliance-oriented strategies often used in conjunction with ESG and CSR(D) reporting and governance. Similarly, funds and value indicate a significant connection to policies, financial instruments, and outcome evaluations in the ESG domain. This uniformity suggests a robust, application-oriented discourse in ESG literature and practices.

Although both datasets address sustainability and entrepreneurship, the INT sample seems to present a more abstract and systemic perspective, which can be suited for conceptual exploration and collaborative frameworks. In contrast, the ESG sample is connected to measurable outcomes, regulatory compliance, and investment rationale. This distinction highlights the need for context-aware ML and NLP strategies when developing tools or insights for diverse audiences, specifically for strategic advisory targeting intermediaries and policy, along with performance enhancement in ESG settings.

*LDAvis*

To enhance the interpretability of the LDA model, we utilized the interactive visualization library pyLDAvis. This tool allows users to explore topics via a two-dimensional representation derived from multidimensional scaling, where the distance between topics indicates their semantic similarities. Each topic is depicted as a circle, with the size representing its overall prevalence in the corpus. The right-side panel offers a detailed view of the most relevant terms for each topic, which are dynamically ranked using a relevance metric that considers both term frequency and exclusivity. We employed the gensimmodels interface of pyLDAvis for seamless integration with the LDA model constructed using Gensim. This visualization method provided essential insights into the identified topics' structure, overlap, and uniqueness, aiding both qualitative interpretation and model enhancement.

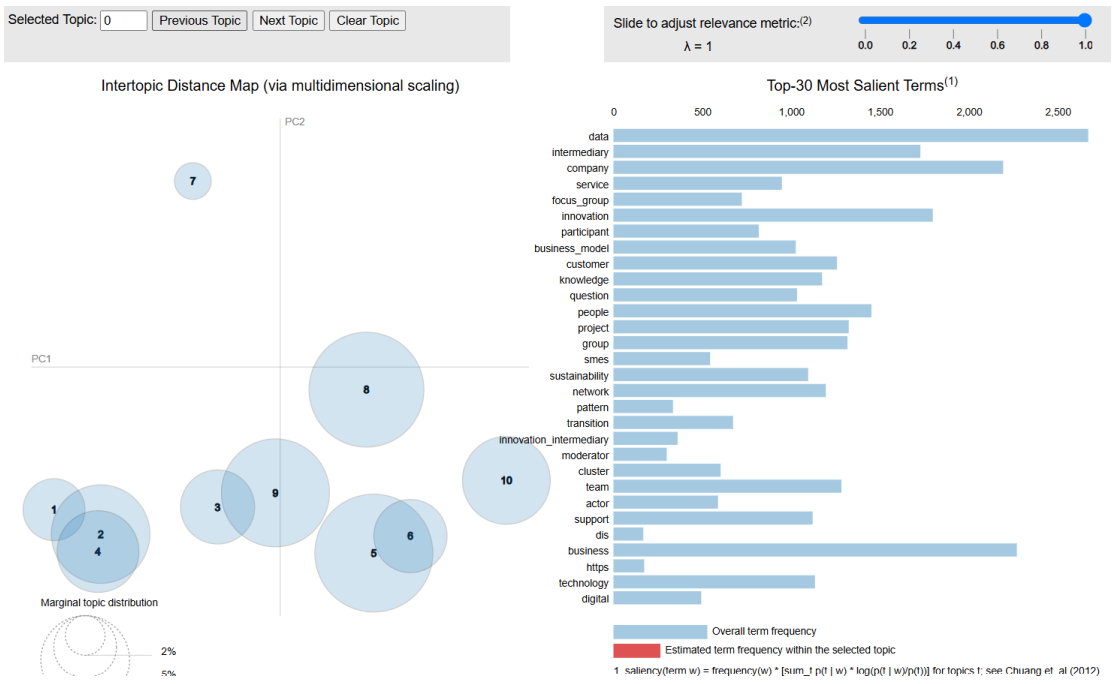


*Figure 4 LDavis Topic Model INT sample*

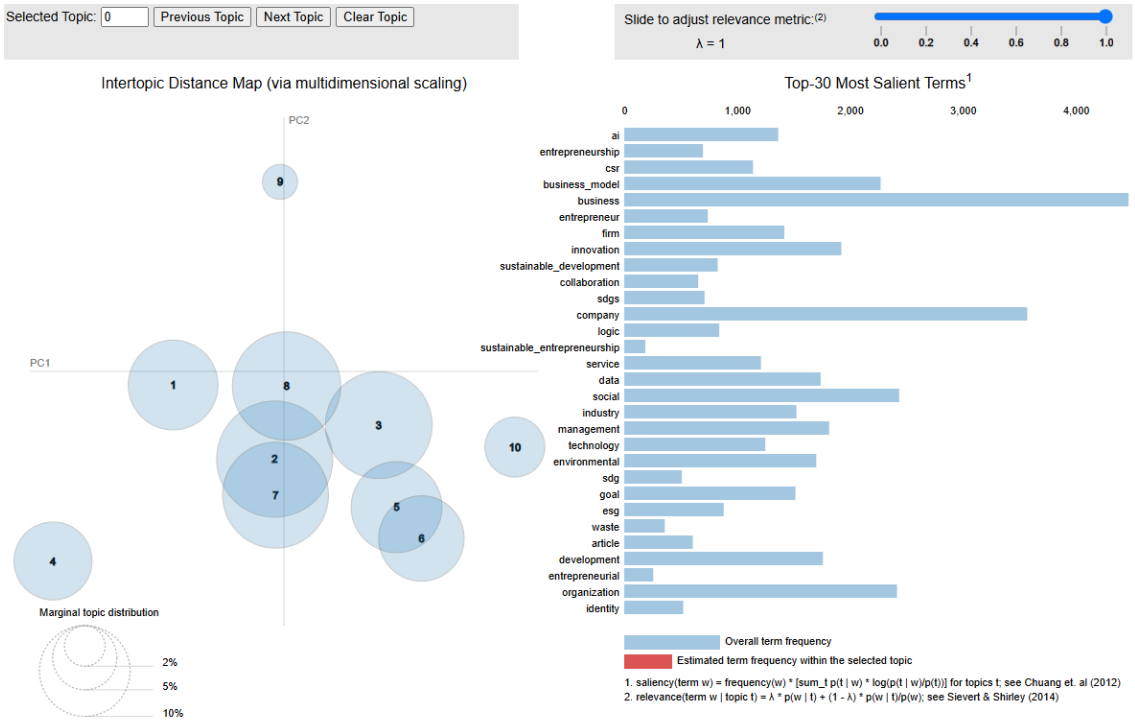


*Figure 5 LDAvis Topic Moded ESG sample*

The LDA visualizations illustrate the thematic structures of the ESG and INT corpora. Within the ESG dataset, the intertopic distance map displays a compact cluster of topics that partially overlap. Key terms such as business model, entrepreneurship, csr, innovation, sustainable development, and data reflect a strong focus on organizational sustainability, corporate responsibility, and data-informed strategies. Moreover, terms such as AI, technology, waste, and ESG underscore an emphasis on digital innovation and environmental impacts in ESG initiatives. The proximity of topics in the ESG model further suggests a closely interlinked discussion centered on practical sustainability and enterprise-led transformations.

In contrast, the INT sample displays a more diverse topic arrangement, featuring several distinct clusters that encapsulate broader conceptual themes. Significant terms such as data, intermediary, focus group, participant, customer, network, and innovation highlight a greater emphasis on stakeholder engagement, participatory research, and innovation ecosystems. Additionally, terms such as team, cluster, and support emphasize collaboration and organizational dynamics, while sustainability appears to be less pronounced in this context. The wider distribution of topics in the INT visualization suggests a more varied discourse, likely reflecting the diversity of contexts and methodologies present in international or interdisciplinary research environments.

These variations in topic distribution and significance underscore the differences between the two datasets: the ESG sample highlights a focused, application-driven narrative on sustainability and corporate performance, while the INT sample offers a wider, more exploratory perspective on stakeholder dynamics and innovation processes.

*ESG-related terminology within samples*

To explore the frequency of ESG-related terminology, we conducted a frequency analysis using a curated list of 17 ESG-related terms drawn from two document samples, one focusing on ESG (the ESG sample) and another from a broader intermediaries (INT) sample (see Figures 6-7). The results indicate a marked difference in term frequency between the two datasets. In the ESG sample, terms such as social, sustainable, and business model were the most common, with each appearing over 5,000 times. In contrast, the INT sample contained much lower frequencies across all terms, with the highest term, social, exceeding just 2,200 mentions.

Both samples displayed some overlap in commonly used terms, particularly social, sustainable, and impact, reflecting a fundamental relevance of ESG discussions even in non-ESG-focused texts. However, the ESG sample emphasized terms like green and business model more heavily, showcasing a stronger focus on sustainability narratives and frameworks. Rarer terms like CSRD, business model tool, and sustainable transition appeared infrequently in both datasets, likely due to their specificity or recent emergence in ESG conversations.

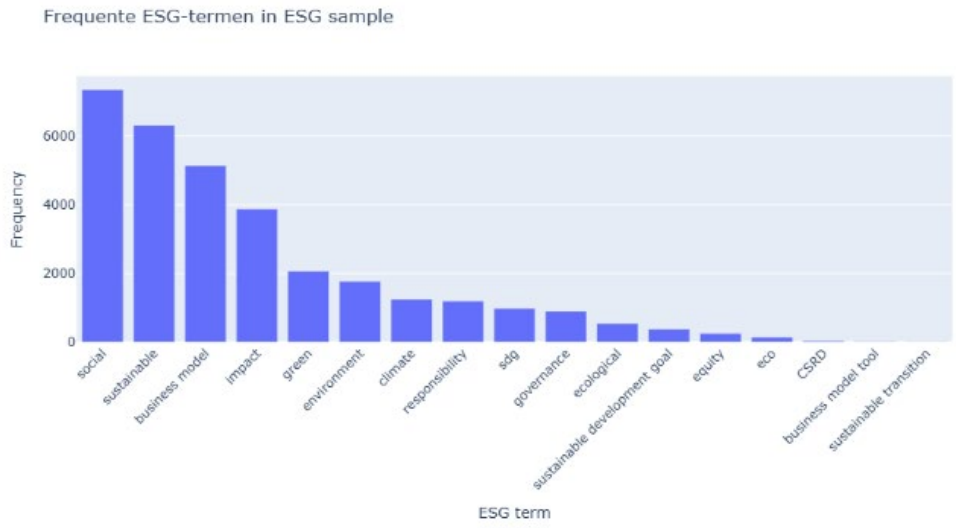


*Figure 6 ESG Term Frequency in ESG Sample*

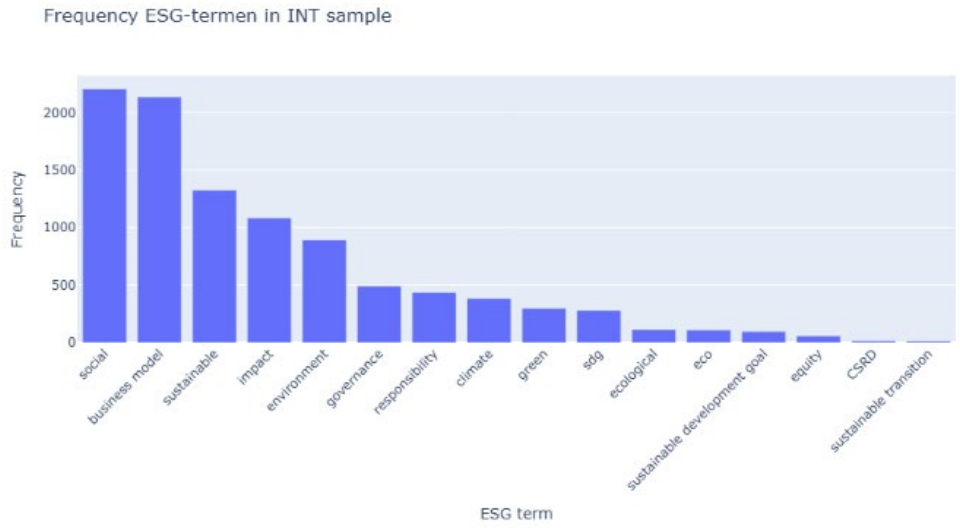


*Figure 7 ESG Term Frequency in INT Sample*

Overall, the findings confirm that ESG-specific terminology is substantially more prevalent in ESG-labeled documents, supporting the effectiveness of the dataset segmentation and highlighting key thematic emphases within the ESG discourse.

During the conference, we will expand upon our preliminary findings, discussions, and conclusions.

## DISCUSSION

The initial results from the group interviews highlight the essential concepts and methodological challenges in creating a context and language-sensitive BMTfS to facilitate transitions in SMEs. One key insight surfaced was the dual role of language: serving both as a means for organizational sense-making and as an analytical tool for ML applications. Participants stressed the importance of carefully managing this duality, especially since interdisciplinary teams have different assumptions concerning language's role and function in advisory contexts.

Language interpretation diverged when it was seen as both a discursive practice and a technical tool. Stakeholders from De Betekenisfabriek emphasized the significance of framing, particularly concerning identity, issues, and roles, as essential for shaping organizational identity and involving stakeholders. Conversely, Eshuis prioritized regulatory frameworks like the Corporate Sustainability Reporting Directive (CSRD), focusing on language use in compliance and formal communication. This difference emphasized the necessity for a common vocabulary and a clear analytical framework to foster effective interdisciplinary collaboration. Throughout the discussions, terms such as language, tool, and model showed varied interpretations, ranging from abstract concepts to specific machine learning applications. Despite this, a consensus gradually developed around BMT as a hybrid structure that combines linguistic, organizational, and technological aspects.

This mutual understanding clarified the tool's objective: to link organizational narratives to measurable sustainability aspects via interpretable ML outputs. Acknowledging the difference between language as influence (framing, sense-making) and language as method (data modeling, NLP) was crucial for achieving epistemological alignment across various fields. The group

unanimously agreed that the effective execution of this tool relies not just on technical viability but also on its significance to advisory practices.

Participants suggested integrating supervised and unsupervised modeling techniques, such as topic modeling, hierarchical clustering, and semantic similarity mapping. These approaches were regarded as technical tools and ways to validate current framing models and sustainability heuristics. By aligning topic clusters generated inductively with categories established deductively (e.g., CSRD criteria, ESG dimensions, language frames), the project aims to develop a data-driven and theory-based classification system.

Importantly, the research highlighted the importance of validating findings in real-world settings. Although the integration of AI techniques into advisory processes shows potential, their actual effectiveness relies on producing actionable insights, facilitating frame alignment, and enhancing strategic decision-making for both clients and advisors. Prioritizing contextual sensitivity and conducting live tests within organizational contexts are vital next steps for tool advancement.

The project gained depth from ML and NLP methods, providing insights into ESG discourse using a subsample of documents on ESG-related topics and on documents on intermediaries in sustainable contexts (INT). Topic modeling with BERTopic and LDA highlighted distinct differences between two datasets: an ESG-centric collection and a larger INT sample. The ESG dataset exhibited closely clustered themes focused on corporate responsibility, ESG investing, circular economy practices, and sustainability metrics. In contrast, the INT sample presented a wider range of exploratory topics that underscored participatory methods, stakeholder dynamics, and innovation ecosystems. These trends were reflected in the pyLDAvis visualizations, which displayed a more compact topic space for the ESG model, compared to the wider configuration of the INT sample.

A complimentary term frequency analysis has visualized the prevalence of sustainability language in the ESG-labeled documents. Although certain words like social, sustainable, and impact were present in both sets, the ESG corpus exhibited notably greater relative and absolute frequencies, especially for terms such as green, business model, and CSR. These quantitative results reinforce the model outputs and substantiate the corpus segmentation employed in this study.

Collectively, the preliminary findings highlight the promise of merging language framing with ML and NLP methods in facilitating sustainability transitions using BMT for SMEs. Shared language creation provides a framework for understanding how organizations create meaning, while ML and NLP techniques enhance scalability, organization, and pattern detection in complex datasets. These methods offer more context-sensitive, actionable tools that connect qualitative insights with data-driven analysis(Messmer et al. 2024). Developing these hybrid tools advances sustainability advisory practices and wider dialogues on how language and machine learning can collaboratively foster social and ecological change.

**Limitations and Future Research**

Our analysis has only recently begun, and we are currently refining the model and the parameters. Therefore, the preliminary findings are mainly descriptive and should not be interpreted as generalizable. Moreover, we plan to conduct a comparative analysis of these subfolders to explore the semantic differences related to context and contrast them with the comprehensive folder that includes all documents. Furthermore, the next step involves thoroughly examining these results and

their comparison with existing scientific literature to gain a deeper understanding, as they currently act only as reflections.

## CONCLUSION

This study's preliminary findings highlight the importance of combining shared language construction, contextual understanding, and the use of ML and NLP techniques in enhancing advisory practices through BMTfS. The results indicate essential epistemological differences regarding language interpretation and usage within organizations, emphasizing the need for unified frameworks to address these disparities. By employing group interviews, topic modeling, and term frequency analysis, this research lays empirical and theoretical groundwork for a BMT that merges linguistic insights with machine learning capabilities.

The comparative analysis of ESG and INT corpora reveals that ESG-focused texts demonstrate greater thematic cohesion and alignment with regulatory and operational frameworks. In contrast, INT texts present a broader, more exploratory discussion. This indicates that language-centric ML can uncover domain-specific insights and facilitate reframing processes in advisory roles, provided these models are trained on high-quality, contextually rich data. It is important to acknowledge the extent to which the selection criteria impact the outcomes and subsequently influence the results.

We intend to enhance our datasets, improve classification models, and test prototypes in real-world advisory settings as we move forward. Additionally, prioritizing interpretability and user adoption is essential to ensure that ML/NLP outputs are comprehensible and actionable for practitioners. In the end, this research contributes to the expanding field dedicated to leveraging language for sustainable transitions, backed by data-driven and conceptually sound tools.

## REFERENCES

Aagaard, A. (2024), *Business Model Innovation: Game Changers and Contemporary Issues*.

Aagaard, A. and C. Tucci (2024), 'AI-Driven Business Model Innovation: Pioneering New Frontiers in Value Creation', in *Business Model Innovation: Game Changers and Contemporary Issues*, Springer International Publishing Cham, pp. 295–328.

Amit, R. and C. Zott (2012), 'Creating value through business model innovation', *MIT Sloan Management Review*.

Asmussen, C. B. and C. Møller (2019), 'Smart literature review: a practical topic modelling approach to exploratory literature review', *Journal of Big Data*, 6 (1), 1–18.

Baker, T. and F. Welter (2020), *Contextualizing Entrepreneurship Theory*.

Baron, R. A. (2006), 'Opportunity recognition as pattern recognition: How entrepreneurs connect the dots to identify new business opportunities', *The Academy of Management Perspectives*, 20 (1), 104–19.

Bera, P., A. Burton-Jones and Y. Wand (2014), 'Research note—how semantics and pragmatics interact in understanding conceptual models', *Information Systems Research*, 25 (2), 401–19.

Bhatnagar, R., D. Keskin, A. Kirkels, A. G. L. Romme and J. C. C. M. Huijben (2022), 'Design principles for sustainability assessments in the business model innovation process', *Journal of Cleaner Production*, 377, 134313.

Breuer, H., K. Fichter, F. Lüdeke-Freund and I. Tiemann (2018), 'Sustainability-oriented business model development: Principles, criteria and tools', *International Journal of Entrepreneurial Venturing*, 10 (2), 256–86.

Burns, D., J. Howard and S. M. Ospina (2021), *The SAGE Handbook of Participatory Research and Inquiry*, SAGE.

Carlsson, R. and T. Nevzorova (2025), 'Measuring sustainable transformation of small and medium-sized enterprises using management systems standards', *Business Strategy and the Environment*, 34 (1), 708–23.

Chauhan, U. and A. Shah (2021), 'Topic Modeling Using Latent Dirichlet allocation: A Survey', *ACM Comput. Surv.*, 54 (7), Article 145.

Creswell, J. W. (2003), 'Research Design: Qualitative, Quantitative, and Mixed Methods Approaches', vol. null.

Divyashree, S., C. J. Joshua, A. Q. Md, S. Mohan, A. S. Abdullah, U. H. Mohamad, N. Innab and A. Ahmadian (2024), 'Enabling business sustainability for stock market data using machine learning and deep learning approaches', *Annals of Operations Research*, 342 (1), 287–322.

Duriau, V. J., R. K. Reger and M. D. Pfarrer (2007), 'A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements', *Organizational Research Methods*, 10 (1), 5–34.

Emerson, J. (2003), 'The Blended Value Proposition: Integrating Social and Financial Returns', *California Management Review*, 45 (4), 35–51.

Evans, S., D. Vladimirova, M. Holgado, K. Van Fossen, M. Yang, E. A. Silva and C. Y. Barlow (2017), 'Business model innovation for sustainability: Towards a unified perspective for creation of sustainable business models', *Business Strategy and the Environment*, 26 (5), 597–608.

Foss, N. J. and T. Saebi (2017), 'Fifteen years of research on business model innovation: How far have we come, and where should we go?', *Journal of Management*, 43 (1), 200–27.

Gamlath, M., C. Gunathilaka, A. Wijesinghe, S. Ahangama, I. Perera and L. Sivaneasharajah (2023), 'An Integrated Approach to ESG Index Construction with Machine Learning', in *2023 Moratuwa Engineering Research Conference (MERCon)*, IEEE, pp. 252–7.

Geissdoerfer, M., M. P. P. Pieroni, D. C. A. Pigosso and K. Soufani (2020), 'Circular business models: A review', *Journal of Cleaner Production*, 277, 123741.

Geissdoerfer, M., D. Vladimirova and S. Evans (2018), 'Sustainable business model innovation: A review', *Journal of Cleaner Production*, 198, 401–16.

Gerring, J. (2007), *Case Study Research: Principles and Practices*, New York: Cambridge University Press.

Grimmer, J., M. E. Roberts and B. M. Stewart (2022), *Text as Data: A New Framework for Machine Learning and the Social Sciences*, Princeton University Press.

Haynes, P. and D. Alemna (2022), 'A systematic literature review of the impact of complexity theory on applied economics', *Economies*, 10 (8), 192.

Holmes, J. and N. Wilson (2022), *An Introduction to Sociolinguistics*, Routledge.

Iturrioz, C., C. Aragón and L. Narvaiza (2015), 'How to foster shared innovation within SMEs' networks: Social capital and the role of intermediaries', *European Management Journal*, 33 (2), 104–15.

Joyce, A. and R. L. Paquin (2016), 'The triple layered business model canvas: A tool to design more sustainable business models', *Journal of Cleaner Production*, 135, 1474–86.

Kanda, W., O. Hjelm, J. Clausen and D. Bienkowska (2018), 'Roles of intermediaries in supporting eco-innovation', *Journal of Cleaner Production*, 205, 1006–16.

Kanda, W., M. Kuisma, P. Kivimaa and O. Hjelm (2020), 'Conceptualising the systemic activities of intermediaries in sustainability transitions', *Environmental Innovation and Societal Transitions*, 36, 449–65.

Kang, Y., Z. Cai, C.-W. Tan, Q. Huang and H. Liu (2020), 'Natural language processing (NLP) in management research: A literature review', *Journal of Management Analytics*, 7 (2), 139–72.

Khoja, F., A. Jeffery, K. Ralph and M. and Yegiyan (2022), 'How SMEs benefit from environmental sustainability strategies and practices', *Supply Chain Forum: An International Journal*, 23 (2), 97–112.

Kivimaa, P., A. Bergek, K. Matschoss and H. van Lente (2020), 'Intermediaries in accelerating transitions: Introduction to the special issue', *Environmental Innovation and Societal Transitions*, 36, 372–7.

Kivimaa, P., W. Boon, S. Hyysalo and L. Klerkx (2019a), 'Towards a typology of intermediaries in sustainability transitions: A systematic review and a research agenda', *Research Policy*, 48 (4), 1062–75.

Kivimaa, P., S. Hyysalo, W. Boon, L. Klerkx, M. Martiskainen and J. Schot (2019b), 'Passing the baton: How intermediaries advance sustainability transitions in different phases', *Environmental Innovation and Societal Transitions*, 31, 110–25.

Koers-Stuiver, D. M., M. B. J. de Lat and T. I. Haaker (2024), *Navigating Sustainable Business Model Transitions Through Tooling, Language, and Context by Intermediaries: A Data Science-Based Longitudinal Case Study*.

Koers-Stuiver, D. M., M. A. A. Kamm and K. Schöller (2023), *Uncovering the Meaning of New Business Models and Multiple Value Creation*.

Köhler, J., F. W. Geels, F. Kern, J. Markard, E. Onsongo, A. Wieczorek, F. Alkemade, F. Avelino, A. Bergek, F. Boons, L. Fünfschilling, D. Hess, G. Holtz, S. Hyysalo, K. Jenkins, P. Kivimaa, M. Martiskainen, A. McMeekin, M. S. Mühlemeier, B. Nykvist, B. Pel, R. Raven, H. Rohracher, B. Sandén, J. Schot, B. Sovacool, B. Turnheim, D. Welch and P. Wells (2019), 'An agenda for sustainability transitions research: State of the art and future directions', *Environmental Innovation and Societal Transitions*, 31, 1–32.

Kundurpi, A., L. Westman, C. Luederitz, S. Burch and A. Mercado (2021), 'Navigating between adaptation and transformation: How intermediaries support businesses in sustainability transitions', *Journal of Cleaner Production*, 283, 125366.

Lecomte, P., M. Vigier, C. Gaibrois and B. Beeler (2023), *Understanding the Dynamics of Language and Multilingualism in Professional Contexts: Advances in Language-Sensitive Management Research*, Edward Elgar Publishing.

Lobo, R. C. G., M. B. Pigatto, D. Denes and A. J. Isaak (2025), 'The Role of Intermediaries in Sociotechnical Transitions: A Systematic Framework Towards Sustainability Management', *Business Strategy and the Environment*, n/a (n/a), accessed at https://doi.org/https://doi.org/10.1002/bse.4237.

Lord, F. N., J. Prior and M. Retamal (2025), 'What factors enable or limit the role of intermediaries in strengthening transformative capacities? Case studies of intermediaries in two Cambodian cities', *Urban Transformations*, 7 (1), 1.

Massa, L., C. L. Tucci and A. Afuah (2017), 'A critical assessment of business model research', *Academy of Management Annals*, 11 (1), 73–104.

Medeiros, M. M. de, N. Hoppen and A. C. G. Maçada (2020), 'Data science for business: benefits, challenges and opportunities', *The Bottom Line*, 33 (2), 149–63.

Messmer, L.-M., C. Reich and D. O. Abdeslam (2024), 'Context-Aware Machine Learning: A Survey', in *Proceedings of the Future Technologies Conference*, Springer, pp. 252–72.

Midgley, R. and S. Arya (2022), 'The Triple Layered Business Model Canvas: A Tool for Sustainability', *International Journal of Management Research and Social Science (IJMRSS)*, 9 (1), 8–12.

Mignon, I. and W. Kanda (2018), 'A typology of intermediary organizations and their impact on sustainability transition policies', *Environmental Innovation and Societal Transitions*, 29, 100–13.

Milei, P., N. Votintseva and A. Barajas (2025), 'Automated Identification of Business Models', *Information Processing & Management*, 62 (1), 103893.

Moghrabi, I. A. R., S. A. Bhat, P. Szczuko, R. A. AlKhaled and M. A. Dar (2023), 'Digital transformation and its influence on sustainable manufacturing and business practices', *Sustainability*, 15 (4), 3010.

Montejo-Ráez, A. and S. M. Jiménez-Zafra (2022), *Current Approaches and Applications in Natural Language Processing*, Basel: MDPI Books.

Moursellas, A., D. De, T. Wurzer, A. Skouloudis, G. Reiner, A. Chaudhuri, T. Manousidis, C. Malesios, K. Evangelinos and P. K. Dey (2023), 'Sustainability practices and performance in European small-and-medium enterprises: Insights from multiple case studies', *Circular Economy and Sustainability*, 3 (2), 835–60.

Munn, Z., M. D. J. Peters, C. Stern, C. Tufanaru, A. McArthur and E. Aromataris (2018), 'Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach', *BMC Medical Research Methodology*, 18 (1), 1–7.

Murto, P., S. Hyysalo, J. K. Juntunen and M. Jalas (2020), 'Capturing the micro-level of intermediation in transitions: Comparing ethnographic and interview methods', *Environmental Innovation and Societal Transitions*, 36, 406–17.

Osterwalder, A. and Y. Pigneur (2010), *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, John Wiley & Sons.

Passaro, R., I. Quinto, G. Scandurra and A. Thomas (2023), 'The drivers of eco-innovations in small and medium-sized enterprises: A systematic literature review and research directions', *Business Strategy and the Environment*, 32 (4), 1432–50.

Pinkse, J., F. Lüdeke-Freund, O. Laasch, Y. Snihur and R. Bohnsack (2023), 'The organizational dynamics of business models for sustainability: Discursive and cognitive pathways for change'.

Quartey, S. H. and O. Oguntoye (2020), 'Promoting corporate sustainability in small and medium-sized enterprises: Key determinants of intermediary performance in Africa', *Business Strategy and the Environment*, 29 (3), 1160–72.

Ritala, P., L. Albareda and N. Bocken (2021), 'Value creation and appropriation in economic, social, and environmental domains: Recognizing and resolving the institutionalized asymmetries', *Journal of Cleaner Production*, 290, 125796.

Salvador, R., P. V. Søberg, M. S. Jørgensen, L.-L. Schmidt-Kallesøe and S. B. Larsen (2023), 'Explaining sustainability performance and maturity in SMEs – Learnings from a 100-participant sustainability innovation project', *Journal of Cleaner Production*, 419, 138248.

Sancak, I. E. (2023), 'Change management in sustainability transformation: A model for business organizations', *Journal of Environmental Management*, 330, 117165.

Sanden, G. R. (2016), 'Language: the sharpest tool in the business strategy toolbox', *Corporate Communications: An International Journal*, 21 (3), 274–88.

Singh, J. (2023), *Natural Language Processing in the Real World: Text Processing, Analytics, and Classification*, Chapman and Hall/CRC.

Spivey, M., K. McRae, M. Joanisse, M. Joanisse, K. McRae and M. Spivey (2012), 'The Cambridge Handbook of Psycholinguistics', Cambridge: Cambridge University Press.

Szopinski, D., L. Massa, T. John, D. Kundisch and C. L. Tucci (2022), 'Modeling business models: a cross-disciplinary analysis of business model modeling languages and directions for future research', *Communications of the Association for Information Systems*, 51 (1), 39.

Tashakkori, A. and C. Teddlie (1998), *Mixed Methodology: Combining Qualitative and Quantitative Approaches*, vol. 46, Sage.

Teece, D. J. (2010), 'Business models, business strategy and innovation', *Long Range Planning*, 43 (2–3), 172–94.

Telukdarie, A., H. Mahure and M. Sishi (2024), 'The digitization of ESG matrix', *Procedia Computer Science*, 239, 808–15.

Tricco, A. C., E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. J. Peters, T. Horsley and L. Weeks (2018), 'PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation', *Annals of Internal Medicine*, 169 (7), 467–73.

van Lente, H., W. P. C. Boon and L. Klerkx (2020), 'Positioning of systemic intermediaries in sustainability transitions: Between storylines and speech acts', *Environmental Innovation and Societal Transitions*, 36, 485–97.

Velter, M. G. E., V. Bitzer and N. M. P. Bocken (2022), 'A boundary tool for multi-stakeholder sustainable business model innovation', *Circular Economy and Sustainability*, 2 (2), 401–31.

Wen, X. and J. R. Taylor (2021), *The Routledge Handbook of Cognitive Linguistics*, Routledge London; New York, NY.

Wilkinson, S. (1998), 'Focus group methodology: a review', *International Journal of Social Research Methodology*, 1 (3), 181–203.