Journal of Business Research 170 (2024) 114254

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Machine learning in marketing: Recent progress and future research directions

Editorial for the Special Issue "Machine Learning in Marketing"

Dennis Herhausen [a,*], Stefan F. Bernritter [b], Eric W.T. Ngai [c], Ajay Kumar [d], Dursun Delen [e,f]

[a] *Marketing, Vrije Universiteit Amsterdam, the Netherlands*
[b] *Marketing, King's Business School, King's College London, United Kingdom*
[c] *Information Systems and Operations Management, The Hong Kong Polytechnic University, Hong Kong*
[d] *Operations, Data & Artificial Intelligence, EM Lyon Business School, France*
[e] *Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Stillwater 74078, OK, USA*
[f] *Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Istinye University, Sariyer/Istanbul, Turkey*

ARTICLE INFO

Keywords:
Machine learning
Privacy
Algorithm
Marketing
Research agenda

ABSTRACT

Decision-making in marketing has changed dramatically in the past decade. Companies increasingly use algorithms to generate predictions for marketing decisions, such as which consumers to target with which offers. Such algorithmic decision-making promises to make marketing more intelligent, efficient, consumer-friendly, and, ultimately, more effective. Not surprisingly, machine learning is a trending topic for marketing researchers and practitioners. However, machine learning also introduces important challenges to the marketing landscape. We discuss this development by outlining recent progress and future research directions of machine learning in marketing. Specifically, we provide an overview of typical machine learning applications in marketing and present a guiding framework. We position the articles in the *Journal of Business Research's* Special Issue on "Machine Learning in Marketing" within this framework and conclude by putting forward a research agenda to further guide future research in this area.

1. Introduction

Marketing decision-makers today often struggle to adequately capture and transform (big) customer data into meaningful insights (e.g., De Luca, Herhausen, Troilo, & Rossi, 2021; Sheth & Kellstadt, 2021). Recent research indicates that machine learning (ML)—a field of computer science dedicated to developing learning algorithms, often using big data, to generate predictions needed to make decisions (Agrawal, Gans, & Goldfarb, 2018)—can help companies manage the flood of data (e.g., Davenport, Guha, Grewal, & Bressgott, 2020; Hagen et al., 2020; Ma & Sun, 2020; Vermeer, Araujo, Bernritter, & van Noort, 2019). ML has been a trending topic in many industries for quite a while now, the marketing industry will be no exception, and it is being used in various industries in the context of both B2C and B2B (e.g., Herhausen, Miočević, Morgan, & Kleijnen, 2020; Kumar, Shankar, & Aljohani, 2020; Luo, Tong, Fang, & Qu, 2019; Rust, 2020). ML promises to make marketing more intelligent, efficient, consumer-friendly, and, ultimately, more effective (Huang & Rust, 2021). To put it more directly, proficiency in ML could become an essential skill for numerous marketing researchers and practitioners rather than just a desirable one.

Faced with this development, we aim to highlight recent progress and future research directions on ML in marketing. Such an endeavor is both important and timely, given the dramatic increase in publications on ML in marketing in recent years, as summarized in Fig. 1. Indeed, from 2012 to 2022, the number of publications has increased by almost 600 percent to more than 50,000 yearly publications. Rather than provide an exhaustive literature review, our purpose here is to discuss the most fundamental concepts and topics from past and present research that will drive future research on ML in marketing.[1] In doing so, we position the articles of this special issue as an overarching framework

* Corresponding author.
E-mail addresses: dennis.herhausen@vu.nl (D. Herhausen), stefan.bernritter@kcl.ac.uk (S.F. Bernritter), eric.ngai@polyu.edu.hk (E.W.T. Ngai), akumar@em-lyon.com (A. Kumar), dursun.delen@okstate.edu (D. Delen).

[1] Ma and Sun (2020) and Ngai and Wu (2022) provide more comprehensive reviews of ML in marketing.

https://doi.org/10.1016/j.jbusres.2023.114254

Available online 7 October 2023
0148-2963/© 2023 Elsevier Inc. All rights reserved.

that touches upon three key themes, namely employing ML in marketing, benchmarking ML in marketing, and managing ML in marketing.

2. Machine learning in marketing

Very broadly, ML can be described as an algorithm that learns from experience concerning some class of tasks and performance measures because its performance at the tasks at hand improves with experience (Mitchell, 1997). The ability of ML to look for patterns in data and enable better decision-making has attracted researchers and practitioners alike, such that it has been widely applied in a variety of business functions, including marketing. Within marketing, ML is a powerful tool for data analysis; it automates analytical model building and can be used for mining large sets of data, providing marketers with opportunities to gain new insights into consumer behavior and improve the performance of marketing operations.

ML in marketing is a vast and rapidly evolving field, encompassing various methods for addressing diverse tasks (Ma & Sun, 2020). Despite this variety, the typical application logic of ML in marketing can be summarized in four steps, as displayed in Fig. 2:

1. Relevant marketing phenomena generate data;
2. This data is preprocessed and annotated, and features are extracted;
3. The enriched data trains and validates an ML model;
4. Classifications and predictions from the ML model inform marketing decisions.

For example, Google generates data from the search behavior of all its users, extracting important features from each data profile. It uses these features to train a recommendation algorithm, whose predictions then inform individual search results and the type of ads displayed to the user. Notably, there is also a feedback loop, such that individual search results and the type of ads displayed influence future search behavior. Building upon this logic, a diverse set of ML methods, such as support-vector machines (Cui & Curry, 2005), topic models (Tirunillai & Tellis, 2014), ensemble trees (Yoganarasimhan, 2020), causal forests (Zhang & Luo, 2023), double ML (Gordon et al., 2023), and deep neural networks (Liu, Dzyabura, & Mizik, 2020) have been used by practitioners and researchers to inform marketing decisions.

Ngai and Wu (2022) conducted a systematic literature review of ML applications in marketing using the 7Ps marketing mix framework from Boom and Bitner (1981). This review shed light on the application focus

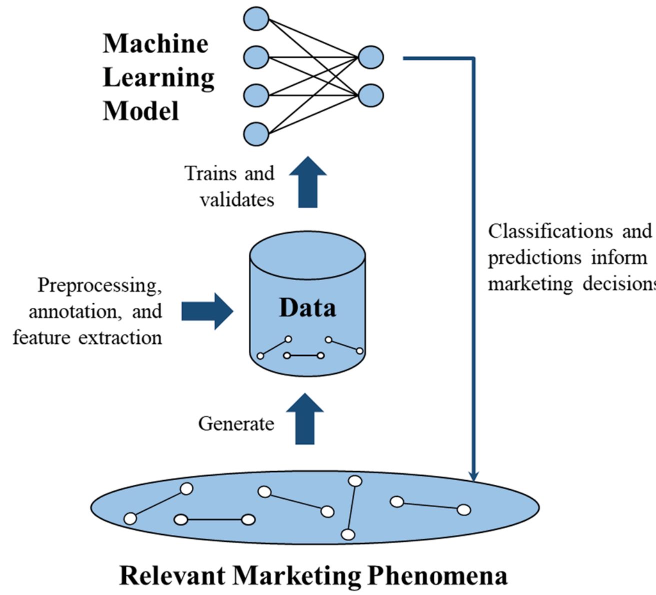

Fig. 2. A Simplified Framework for Machine Learning in Marketing. Note: Inspired by van Giffen et al. (2022).

of this research stream (see Fig. 3). Out of 140 identified articles published up to 2021:

- 32% examined people-related marketing activities, including churn prediction, targeting customer prediction, engagement, and facial recognition.
- 24% examined promotion-related marketing activities, including advertising management, demand prediction, and chatbots.
- 23% examined product-related marketing activities, including product recommendations, brand and trademark management, and purchase decision prediction.
- 10% examined process-related marketing activities, such as market or customer segmentation.
- 5% examined physical evidence-related marketing activities, such as content about the environment in which the service is delivered and tangible information about physical goods.

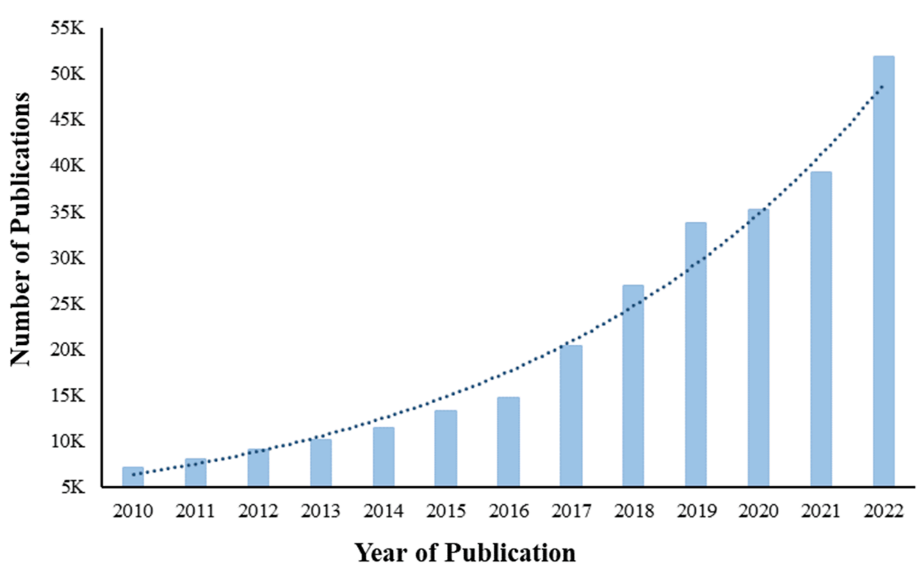

Fig. 1. Publications on Machine Learning in Marketing over Time. Note: We searched for publications that include “machine learning” and “marketing” on google scholar for the respective years and report the number of publications per year in thousands.

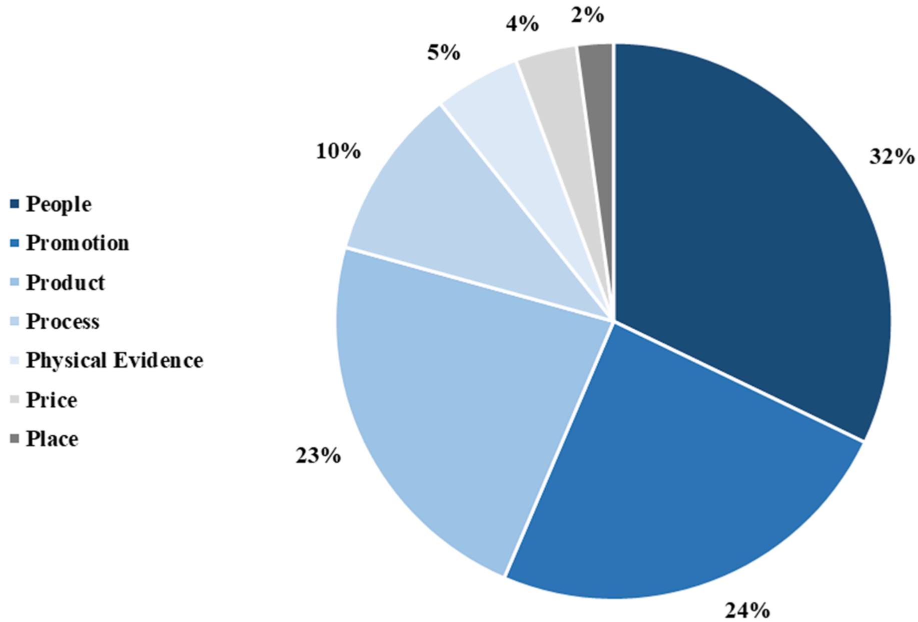

Fig. 3. The focus of ML Applications in the Marketing Literature. Note: We thank Ngai and Wu (2022) for sharing their data on ML applications in marketing with us. Their literature review is based on a search query on October 1, 2021, with the following retrieval string in SCIE and SSCI: TS = ("machine learning" OR "neural network" OR "artificial intelligence") AND TS = ("marketing" OR "retailing"). A total of 962 articles were initially retrieved, and 140 articles were eligible for the categorization.

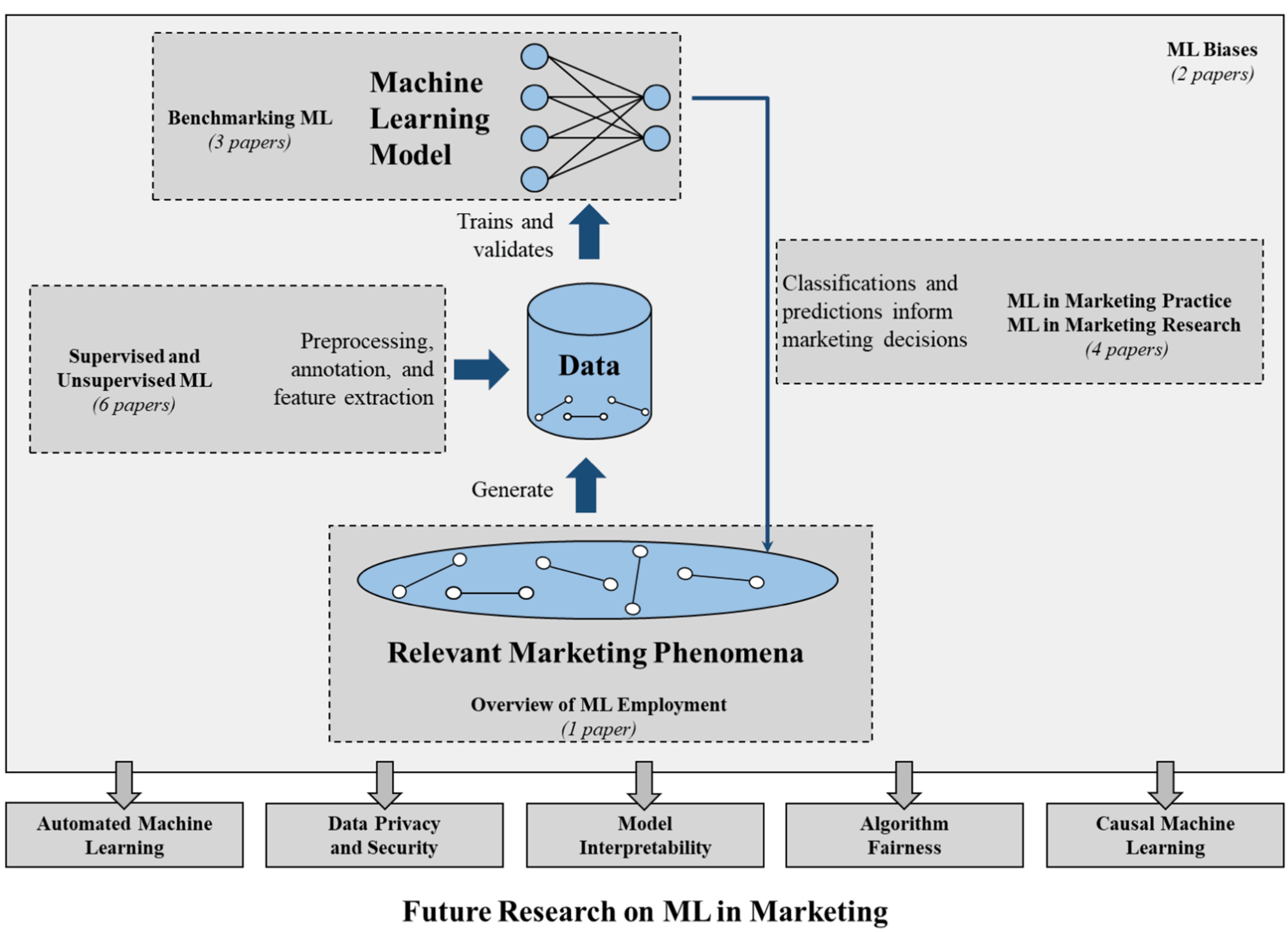

Fig. 4. Key Areas of Articles in the Special Issue and Future Research Directions.

- 4% examined price-related marketing activities, such as the effects of pricing structures and price forecasting.
- 2% examined place-related marketing activities, such as multi-channel management, and the effect of geographical proximity.

However, we see two essential areas for improvement in many of these studies. First, it must be noted that marketing relationships uncovered using ML are often correlational rather than causal (for an excellent discussion, see Langen & Huber, 2023). Little attention has been paid to endogeneity concerns, even when a predictive focus is used. Issues such as selection, omitted variables, and simultaneity, addressed in econometric models, are typically ignored in ML algorithms. Second, many ML algorithms are black boxes that can often accurately form predictions. However, they do not permit marketing researchers and managers to understand the causes or reasons for a decision. Algorithms often lack interpretability due to their opaque model structure and the absence of clear connections between variables. Thus, causal ML and model interpretability are among the important future research directions on ML in marketing that we will discuss later.

Moreover, ML algorithms differ substantially from traditional data analysis methods used in the past for decision support in marketing. ML algorithms follow a probabilistic approach in which decisions are made by learning patterns from training data and applying these decisions to new data. Decision support from ML algorithms is provided in the form of probabilities, leading to different levels of uncertainty in the results obtained. Despite these challenges, the recent progress in understanding, applying, and managing ML in marketing is impressive.

3. Recent progress from the Special Issue

We are pleased to present this Special Issue of the *Journal of Business Research* on "Machine Learning in Marketing." We aimed to examine the current and future impact of ML and related technologies in marketing. The Special Issue contains 16 papers selected from 75 submissions drawn from a wide range of topics, methodologies, and perspectives. We have grouped the articles into three broader themes (i.e., employing ML in marketing, benchmarking ML in marketing, and managing ML in marketing) and seven sub-themes (i.e., an overview of ML employment, supervised ML, unsupervised ML, benchmarking ML, ML biases, ML in marketing practice, and ML in marketing research), as summarized in Fig. 4. Next, we summarize the key insights from each article of this special issue and embed these insights into the respective theme. The order in which the articles are presented is based on the theme studied, as we have grouped articles around the three broader themes (see Table 1).

3.1. Employing machine learning in marketing

The initial theme focuses on the utilization of ML in marketing. The articles included in this theme showcase innovative ML techniques that address marketing challenges or examine the ways in which ML is applied in marketing. Seven articles of the Special Issue address open questions regarding the employment of ML in marketing. Ngai and Wu (2022) use a systematic literature review of 140 articles to provide an overview of ML employment in marketing. The authors develop a framework that summarizes ML applications in marketing based on the 7Ps marketing mix framework (i.e., product, price, promotion, place, people, process, and physical evidence) and the primary ML tools used in marketing such as text, voice, image, and video analytics as well as major ML technologies used in marketing such as supervised, unsupervised, and reinforcement learning algorithms. Based on their findings, the authors propose a two-layer conceptual ML-based marketing application framework that puts forward an agenda for future research.

Three articles use supervised ML to shed light on essential marketing outcomes. Lutz, Pröllochs, and Neumann (2022) build upon information overload theory and implement a state-of-the-art multi-instance

Table 1
Overview of Articles in the Special Issue.

Article	Author(s)	Key Focus	Methodology	Key Insights
Employing ML in Marketing				
1.	Ngai and Wu (2022)	Overview of ML employment	Systematic literature review	The authors develop a framework that summarizes ML applications in marketing based on the 7Ps marketing mix framework and the major ML tools and technologies used in marketing.
2.	Lutz et al. (2022)	Supervised ML	Bidirectional encoder representations from transformers	The authors use a multi-instances learning approach to study the line of argumentation in of Amazon customer reviews at the sentence-level.
3.	Philp et al. (2022)	Supervised ML	Image classification and experiment	The authors use Google Vision AI on restaurants' Instagram posts to analyze how the visual characteristics of product offerings relate to social media engagement.
4.	Sengupta et al. (2021)	Supervised ML	Text mining and hurdle models	The authors use a text mining framework and hurdle models to examine the predictors of successful Airbnb bookings.
5.	Kolomoyets and Dickinger (2023)	Supervised and unsupervised ML	Structural topic modeling and sentiment analyses	The authors employ structural topic modeling and sentiment analyses to provide an empirical account of value alignment from the perspective of customers and service providers.
6.	Airani and Karande (2022)	Unsupervised ML	Bayesian network analysis	The authors use Bayesian network analysis to investigate how consumer sentiments on social media are shaped by user anonymity and authority.
7.	Liu et al. (2022)	Unsupervised ML	Latent Dirichlet allocation and structural equation modelling	The authors use latent Dirichlet allocation and structural equation modeling to examine the effect of value perceptions on

(*continued on next page*)

Table 1 (*continued*)

Article	Author(s)	Key Focus	Methodology	Key Insights
				gifting behavior on sport live streaming platforms.
Benchmarking ML in Marketing				
8.	Feng et al. (2022)	Benchmarking ML	Dynamic ensemble selection	The authors use bank telemarketing data to benchmark a dynamic ensemble selection method with several other state-of-the-art ML methods.
9.	Esmeli et al. (2022)	Benchmarking ML	Computational experiments	The authors present a novel framework of early purchase prediction in online sessions and benchmark the performance of different data mining models.
10.	Potrawa and Tetereva (2022)	Benchmarking ML	Convolutional neural networks	The authors enhance the hedonic pricing framework by adding image and text sources to conventional data and by uncovering new pricing factors and complex relationships that could not be captured by conventional models.
Managing ML in Marketing				
11.	van Giffen et al. (2022)	ML biases	Systematic literature review and in-depth case study	The authors identified eight distinct ML biases, summarized these biases in the cross-industry standard process for data mining, and outlined 24 mitigation methods in a real-world case study.
12.	Akter et al. (2022)	ML biases	Systematic literature review and in-depth interviews	The authors identified three primary dimensions (i.e., design, contextual, and application bias) and ten corresponding subdimensions of algorithmic bias in marketing models.
13.	Volkmar et al. (2022)	ML in marketing practice	Delphi study, survey, focus groups	The authors explore the drivers of and barriers to ML in marketing practice in three distinct domains: culture, strategy, and implementation; decision-making and ethics; and customer management.
14.	Latinovic and Chatterjee (2022)	ML in marketing practice	Conceptual paper	The authors develop an organizational process model in which ML facilitates communication, coordination, and customization in customer-centric organizations.
15.	Ghouri et al. (2022)	ML in marketing practice	Semi-structured interviews	The authors propose micro foundations of social media routines that consist of three processes and four stages, and integrate ML to manage user engagement.
16.	Ordenes and Silipo (2021)	ML in marketing research	Visual-based programing with the KNIME Analytics Platform	The authors create a live repository of ML projects, hosted on the KNIME hub, where researchers can learn, share, and reuse workflows of five annotated projects in the areas of customer churn, sentiment analysis, automated image analysis, search engine optimization, and customer experience.

learning approach (i.e., bidirectional encoder representations from transformers) to study the line of argumentation in 61,837 Amazon customer reviews for 4,647 low-involvement products and 2,335 high-involvement products. The authors find the line of argumentation and review length are closely intertwined, so longer reviews with frequent changes between positive and negative arguments are perceived as less helpful. This insight challenges the assumption that longer reviews are uniformly more helpful. Instead, they argue the effect depends on the complexity of the line of argumentation at the sentence level.

Philp, Jacobson, and Pancer (2022) use image classification with Google Vision AI on 10,173 Instagram posts from 871 restaurants in 26 cities to analyze how product offerings' visual characteristics relate to social media engagement. Results demonstrate food images higher on food typicality are positively associated with engagement regarding likes and comments. A follow-up experiment shows that exposure to typical-appearing foods elevates positive affect, suggesting they are easier to mentally process, which drives engagement. Therefore, contrary to conventional social media practices and food industry trends, the more typical a food appears, the more social media engagement it receives.

Sengupta, Biswas, Kumar, Shankar, and Gupta (2021) use a text mining framework with hurdle-based Poisson and negative binomial regressions on Airbnb data for 22 cities on four continents to examine the predictors of room bookings. They found that super host status, host response time, and communication with guests emerged as the most significant predictors. The authors further executed variable-importance algorithms to seek the top predictors for each continent. They found a largely generalizable empirical framework for host-based and guest-based predictors across Europe, Australia, Asia-Pacific, and North America.

Three articles use unsupervised ML to explore important marketing outcomes. Kolomoyets and Dickinger (2023) employ structural topic modeling, sentiment analyses, and a gradient boosting algorithm with regularized model formalization on 17,372 online hotel reviews from Booking.com of 1,390 four and five-star hotels in six European cities, the respective hotel descriptions from the official hotel websites, and 31,389 hotel descriptions and customer reviews from TripAdvisor.com. By doing so, the authors generate insights into the attributes valued by customers and service providers, the valence of those attributes, the sources of value formation, value alignment between customers and service providers, and the relative importance of value attributes for budget and upscale hotels. Their findings guide managers in monitoring the content of reviews relevant to value creation by directing them toward the different attribute combinations of most significant relevance to hotels in various price segments.

Airani and Karande (2022) build upon consumers' social media journey to investigate how consumer sentiments on social media are shaped by individual users' anonymity and authority, employing Bayesian network analysis on over half a million posts for 127 movies on Twitter. Relying on the do-calculus methodology proposed by Pearl (1995), the authors find that anonymity and authority influence platform effects, including hashtag position and positive and negative bandwagon effects, which impact expressed sentiments. Managers can leverage these findings to track anonymity and authority as well as platform effects to predict and influence consumer sentiments on social media.

Liu, Tan, and Pawar (2022) use latent Dirichlet allocation and structural equation modeling on 16,204 real-time messages and 5,540 virtual gifts collected on the final matchday of the International Table Tennis Federation World Tour Grand Final 2019 from a Chinese sports live streaming platform to examine the effect of value perceptions on gifting behavior. Their results reveal hidden topics important to understanding the perceptions of viewing sporting events in the context of social live-streaming services, offering marketers guidance on improving their engagement with viewers on a sports live-streaming platform.

3.2. Comparing the effectiveness of machine learning in marketing with other methods through benchmarking

Three articles of the Special Issue benchmark multiple ML methods in a marketing context. Feng, Yin, Wang, and Dhamotharan (2022) use a dynamic ensemble selection method that considers the accuracy and average profit with meta-training to predict the success of bank telemarketing sales of time deposits. Compared with mainstream ML methods, including single classifiers and ensemble learning classifiers, the predictive performance of the proposed method is superior in terms of ML and economic metrics. In addition, the authors provide a post hoc explanation of the predicted results so that marketing managers can understand the mechanisms and dynamics that lead to such results.

Esmeli, Bader-El-Den, and Abdullahi (2022) use four state-of-the-art ML models, including random forest, decision tree, bagging, and a deep neural network, to predict the conversion rate (i.e., the proportion of visits ending in sales) in online sessions for registered and unregistered consumers as soon as they land on an e-commerce platform with four real-world datasets spanning millions of e-commerce platform visits. For unregistered consumers, random forest showed superiority over all the datasets in original and re-sampling applied settings. In contrast, for registered consumers, the deep neural network determines purchase decisions better than other models, giving managers an indication of when to use which ML model.

Potrawa and Tetereva (2022) employ convolutional neural networks to enhance the hedonic pricing framework by adding image and text sources to conventional data and by uncovering new pricing factors and complex relationships conventional models are unable to capture. Compared to traditional methods such as the linear, spatial, and nonparametric approaches, the proposed alternative combining feature extraction, predictive modeling, and model interpretation leads to a 25% increase in predictive accuracy. This helps managers understand customers' perception of the value of constituent characteristics in the real estate market.

3.3. Managing machine learning in marketing

Two articles of the Special Issue address the critical issue of ML bias. When ML is used for decision-making in marketing, bias rooted in unrepresentative datasets, inadequate models, weak algorithm designs, or human stereotypes can lead to low performance and unfair decisions, resulting in financial, social, and reputational losses. van Giffen, Herhausen, and Fahse (2022) combine a systematic literature review of 68 articles with an in-depth case study to identify eight distinct ML biases: social bias, measurement bias, representation bias, label bias, algorithmic bias, evaluation bias, deployment bias, and feedback bias. The authors summarize these biases in the cross-industry standard process for data mining and outline 24 mitigation methods. These insights caution managers that there is no bias-free ML, no universal panacea for ML biases, and that the dynamic context in which the marketing decision-making takes place triggers ML biases.

Akter et al. (2022) complement the previous study by triangulating a systematic literature review of 25 articles and 25 in-depth interviews with professionals involved in ML-based marketing model development and execution. The authors identified three primary dimensions (i.e., design bias, contextual bias, and application bias) and ten subdimensions of algorithmic bias in marketing models (i.e., model, data, method, cultural, social, personal, product, price, place, and promotion). They further match the primary dimensions with the subdimensions in research propositions. These findings alert managers to pay attention to numerous sources of biases in ML models in marketing.

Three articles in the Special Issue focus on ML in marketing practice. Most companies do not fully exploit ML's potential, particularly in marketing, where its possible use cases extend beyond mere segmentation and personalization. Volkmar, Fischer, and Reinecke (2022) combine a two-round Delphi study based on personal interviews with expert practitioners and academics, a survey of 101 experienced managers working at the intersection of marketing and ML, and two focus groups involving 11 additional marketers with experience in ML with exploring the drivers of and barriers to ML in marketing practice. They find challenges in three domains: culture, strategy, and implementation; decision-making and ethics; and customer management. These insights lead to several research propositions that address the ML challenges marketing managers and organizations face.

In their conceptual contribution, Latinovic and Chatterjee (2022) develop an organizational process model in which ML facilitates communication, coordination, and customization in customer-centric organizations. They clarify the role of ML-based technologies in improving the three value delivery processes and provide examples of ML-powered solutions in B2B sales and marketing, technology, healthcare, and education. Nevertheless, the authors conclude that relationships remain the salient connections that bind organizations to value delivery. Thus, they emphasize the critical role of employees and their ever-shifting interaction with organizational culture in value-added activities for customers – even when those involve ML.

Ghouri, Mani, ul Haq, and Kamble (2022) used 15 semi-structured interviews with the top management of firms that employ ML-based technologies to shed light on social media use in B2B contexts. Based on these interviews, the authors propose a four-stage three-process model, from customer engagement on social media to the appropriate customer-engagement response. In these micro-foundations of social media use, customers are managed individually by ML, and the ML-integrated routine model provides managers the opportunity to establish strong relationships with their customers.

While ML promises great value for marketing research, the proliferation of data types, methods, tools, and programming languages hampers knowledge integration among marketing researchers, making collaboration difficult. Ordenes and Silipo (2021) address this challenge and provide valuable resources for ML in marketing research. Relying on visual-based programming with the KNIME Analytics Platform (https://www.knime.com), the authors create a live repository of projects hosted on the KNIME hub, where users can learn, share and reuse workflows with visual code. They further developed a starting set of five annotated projects in customer churn, sentiment analysis, automated image analysis, search engine optimization, and customer experience. Two of these projects offer a step-by-step guide to facilitate the learning experience.

4. Future research on machine learning in marketing

This Special Issue presents comprehensive insights into recent progress and the state-of-the-art of ML in marketing. The future offers important research opportunities that can build on this extant knowledge, and we outline a series of suggestions that warrant investigation. These future research directions are related to the themes of (1) automated machine learning, (2) data privacy and security, (3) model interpretability, (4) algorithm fairness, and (5) causal machine learning (see Fig. 4). When discussing the future research directions, we also develop some specific research questions, summarized in Table 2.

Table 2
Selected Future Research Directions for Machine Learning in Marketing.

Research Area	Selected Research Questions
Automated machine learning	1. What are the benefits and pitfalls of fully automating marketing analytics activities with automated ML? How might such activity hinder genuine creativity in marketing initiatives? 2. What role will automated ML play in assisting and potentially replacing marketing analytic professionals? 3. What are the additional concerns (or advantages) regarding data security and privacy with employing automated ML for marketing applications?
Data privacy and security	1. What are the ethical and legal implications of using personal data to drive personalized marketing, and how can these implications be addressed through privacy-preserving ML algorithms and other measures? 2. What are the trade-offs between personalization and privacy, and how can these trade-offs be optimized to maximize consumer trust and satisfaction? 3. What best practices can be identified for compliance with GDPR and CCPA in ML-based marketing, and how can these be integrated into marketing strategies?
Model interpretability	1. When and where is model interpretability more important than predictive power in marketing analytics? 2. How critical are transparency and interpretability in marketing applications for customer acceptance? 3. What is the tradeoff between accuracy and interpretability in marketing analytics? Where is the point of compromise?
Algorithm fairness	1. How can marketers mitigate algorithm bias and improve AI fairness for making better marketing strategies? 2. How to develop fair personalization engines that help marketers to deliver hyper-personalized content for improving targeted advertising? 3. What will be the impact of “fairness” on marketing analytics in the near future?
Causal machine learning	1. To what extent can insights generated by ML be replicated using causal ML? 2. How to ensure data quality that allows causal ML to perform more closely compared to the accuracy of randomized controlled trials? 3. What are best practices in causal ML in marketing?

Note: GDPR = General Data Protection Regulation, CCPA = California Consumer Privacy Act.

4.1. Automated machine learning

Automated ML (AutoML) offers methods and processes for making ML available to nonexperts by automating end-to-end processes of applying ML to real-world problems. The term “automation” has become a popular buzzword in the business world, referring to the increasing use of AI and ML to computerize various business functions that humans previously carried out. This includes tasks performed by white-collar employees, such as decision-making regarding hiring and promotion (Kolbjørnsrud et al. 2016). The replacement of human-performed tasks by technology and computers is not new, as has been the case with agricultural, industrial, and informational revolutions. The latest trend in this continuum has been intelligence requiring human workers to be replaced (or augmented) by computers, AI, and ML. After a series of human versus computer competitions, AI and ML have become the potential agents to automate the underlying processes and replace humans in intelligence-requiring tasks (Delen, 2020). For instance, Deep Blue has beaten the grand master of Chess, Gary Kasparov; IBM Watson has beaten the two best players in the television show Jeopardy!; and AlphaGo has beaten the world league champion in the game of Go. With automated ML, AI is now focusing on automating itself (to automate its own development processes), ultimately aiming to replace (at least for some of the time-demanding tasks) the data scientist.

AutoML refers to the process of computational algorithms taking over the tasks concerning the development of the best possible ML model. Developing the best possible ML model for a given task using disparate and multi-faceted data sources is a time-demanding, experimental process. Considering the possible combinations of design choices related to data and the ML model hyperparameters, the problem can be characterized as NP-hard (i.e., all combinations are deemed near infinite).[2] As no optimal recipe or best practice works with every data and problem pair for this NP-hard problem, a data scientist is expected to spend considerable time trying as many combinations and configurations as possible to discover/learn a good enough ML model. AutoML is designed to handle these repetitive, combinatorial, time-demanding, mundane tasks so that data scientists can spend more time on the creativity-requiring, artistic parts of the process.

Although AutoML has been used almost exclusively in supervised learning problems (i.e., classification and regression), where data is labeled so that the labels can be used to guide/supervise the learning and improvement process, it can potentially be used for unsupervised learning ML tasks such as clustering and association, if a quantifiable improvement metric is designed and employed. Regardless of the learning tasks, AutoML can help produce better models for complex and inherently dynamic marketing problems, such as predicting at-risk customers (churn analysis), determining customer lifetime value, improving product bundles (cross-selling), optimizing marketing

[2] In computational complexity theory, NP-hardness (non-deterministic polynomial-time hardness) is the defining property of a class of problems where there is no known way to find an answer quickly, but if one is provided with information showing what the answer is, it is possible to verify the answer quickly. The informal term quickly means the existence of an algorithm solving the task that runs in polynomial time, such that the time to complete the task varies as a polynomial function on the size of the input to the algorithm (as opposed to e.g., exponential time).

campaigns, and improving the efficacy of customer segmentation.

AutoML has its limitations and challenges (Vaccaro, Sansonetti, & Micarelli, 2021). For instance, it is not suitable for "soft" analytics tasks such as analytics problem/opportunity identification (business understanding), relevant data identification, acquisition, and curation (data understanding), and model evaluations and deployment (privacy and ethical perspective, end-user related change management). AutoML is very good at executing relatively straightforward repetitive data manipulation and model development/testing/improvement tasks. For now, tasks that are non-routine, require creativity, and are end-user/business-knowledge oriented are out of AutoML ML's realm of functionality.

4.2. Data privacy and security

Privacy and security of data are gaining momentum in the research community, mainly due to emerging technologies like cloud computing, analytics engines, and social networks. The use of ML further raises concerns regarding data privacy and security, particularly when it comes to sensitive or personal information. For example, ML algorithms often rely on large datasets to learn and improve their accuracy, but these datasets may contain sensitive personal information such as names, addresses, and financial information. If healthcare datasets are breached, it could lead to the exposure of sensitive personal data and medical identity theft. In the following section, we propose several avenues for future research in the domain of ML related to data privacy and security in marketing.

The first avenues relate to ML algorithms and systems. ML algorithms that can analyze data without compromising individuals' privacy can be developed. Privacy-preserving mechanisms are crucial to protect users' information (Cunha, Mendes, & Vilela, 2021). Several privacy-preserving mechanisms have been developed for data protection at different stages (e.g., data generation, storage, and processing) of its life cycle (Mehmood et al., 2016). These mechanisms can involve techniques such as differential privacy (Zhu, Ye, Wang, Zhou, & Yu, 2022) and secure multiparty computation (Bringer, Chabanne, & Patey, 2013). Research focusing on the challenges and opportunities these various approaches bring for companies and consumers would be very much welcome to help develop privacy-preserving ML algorithms. Moreover, security is specified by several aspects, including confidentiality, integrity, availability, authentication, non-repudiation, authorization, and access control (De Cremer & Kasparov, 2022). Future research could focus on how secure ML systems can be designed and developed to prevent unauthorized access, data breaches, and other security threats. It would also be worthwhile to examine consumers' responses to ML security and how the degree of security in ML systems affects consumer decision-making.

Another important research avenue relates to soft biometrics. Soft biometrics i refer to the use of non-intrusive and behavior-based features to describe individuals (Park & Jain, 2010). Unlike traditional biometrics (e.g., fingerprints, iris scans), soft biometrics capture characteristics such as height, weight, gender, clothing, and gait. Using soft biometrics in ML-based marketing can help preserve consumer privacy and security by avoiding the need for invasive or sensitive information. Within this domain, potential avenues for future research include developing soft biometric models. How can soft biometric models be integrated into marketing workflows to provide personalized recommendations and other services while ensuring consumers control their data and its usage? Moreover, it is important to identify potential risks and harms. What are the potential risks and liabilities associated with using soft biometric models in marketing, and how can these risks be mitigated through privacy-preserving measures and other safeguards?

The European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) are the two most prominent legislative changes affecting data privacy and security. Important ethical and legal considerations related to consumer privacy and security must be considered, particularly considering the GDPR and CCPA. One avenue is evaluating the effectiveness of current privacy regulations. The efficacy of current privacy regulations, such as GDPR and CCPA, in protecting individuals' privacy in ML-based marketing must be evaluated. Such research can help identify regulatory gaps and inform future policy development. More specifically, evaluating the effect of GDPR and CCPA on ML-based marketing is also warranted. In particular, while there was e an era where data was emerging in abundance, we now see data "disappearing" in "black holes" by being locked away after privacy walls. This evaluation can involve surveys or interviews with marketing professionals to better understand the challenges and opportunities these regulations present. In addition, real-world marketing campaigns can be analyzed to identify best practices for compliance.

More broadly, future research in ML-based marketing and privacy/security needs to address the complex interplay between technology, psychology, ethics, and regulation to ensure that ML is used responsibly and beneficially, respecting consumer privacy and security.

4.3. Model interpretability

One major obstacle to the widespread adoption of ML is that computers typically do not provide explanations for their predictions. However, being able to interpret the outputs of a model is crucial for evaluating its validity, building user trust, and using the model's insights to improve its predictive accuracy. We believe this is an area where further research will be especially important to allow the models to gain greater traction among marketing practitioners.

In business analytics, specifically predictive analytics, there has been a trade-off between model interpretability and more complexity; the more complex the ML models (e.g., deep neural networks, random forest, and gradient-boosted machines), the more accurate the prediction outcomes become, but in return, they lose interpretability. As the model gets simpler, moving towards parsimony, the less predictive performance is observed, but at the same time, the more interpretable (and transparent) the model becomes. The seminal research paper by Ribeiro, Singh, and Guestrin (2016) entitled "Why Should I Trust You?" appropriately highlights the issue with ML models and their infamous nature of being perceived as black boxes. With new and enhanced interest in ML methods, especially since the advent of model ensembles and deep learning, model interpretability has become a fast-growing field of research.

To mitigate this deficiency (the so-called "black-box syndrome") within the last couple of decades, the ML community proposed several methods. Some of these methods are global (explaining the factors/variables based on the average score of all data samples), and some are local (providing single sample-level explanations). In predictive modeling, sensitivity analysis usually refers to an exclusive experimentation process designed and executed to discover the cause-and-effect relationship between the input and output variables. Some variable importance methods are model/algorithm specific (i.e., applied to decision trees, neural networks, or random forest), and some are model/algorithm agnostic (i.e., applied to every predictive model).

However, marketing research has only just begun to use these methods (notable exceptions are, e.g., Liu, Lee, & Srinivasan, 2019; Wang, He, Jin, & Hu, 2022). Thus, we briefly present the most commonly used variable importance methods employed in ML and predictive modeling. The first approach is a well-trained decision tree model to identify the relative importance of the input variables—the closer a variable is to the root of the tree, the greater the importance/relative contribution of that variable to the prediction model.

The second approach is a rich and large random forest model that assesses the variable split statistics. If the ratio of a given variable's selection as candidate counts (i.e., the number of times a variable was selected as the level-0 splitter divided by the number of times it was picked randomly as one of the split candidates) is larger, then its importance/relative-contribution is also greater. This process can be extended to the top three layers of the trees to generate a weights average of the split statistics and can be used as a measure of variable importance for random forest models.

The third approach is a sensitivity analysis based on input value perturbation, where the input variables are gradually and systematically changed/perturbed one at a time, and the relative change in output is observed—the larger the change in the output, the greater the importance of the perturbed variable. This method is often used in trained feed-forward neural network modeling where all input variables are numeric and standardized/normalized.

The fourth approach is a sensitivity analysis based on a leave-one-out methodology. This method can be applied to any predictive analytics method (i.e., predictive model agnostic). Because of its general applicability and ease of implementation, this sensitivity analysis method is used as default in several commercial and free or open-source analytics tools. In practice, the method follows a simple experimental design where each variable is excluded from the model development and testing process, and the performance degradation of the model in its absence is assessed as compared to the base model's performance (where all variables are included). If the removal of a variable results in a significant decrease in performance, then that variable is considered more significant or plays a greater role in the prediction system. Once repeated for all variables, the relative importance measures can be normalized, tabulated, or charted for a variable importance report.

The fifth approach is sensitivity analysis based on developing a surrogate model to assess the variable importance of a single record or sample using local interpretable model-agnostic (LIME) or Shapely additive explanations (SHAP) methodologies (Gramegna and Giudici 2021).[3] While the previous methods are considered global interpreters, LIME and SHAP are called local because they explain the importance of the variables at the sample level (as opposed to the average of all samples). These are the latest model interpretability techniques that use a two-model structure; one for prediction and another (simpler one) for interpretation.

4.4. Algorithm fairness

Bias and fairness in AI are two sides of the same coin, and the algorithm fairness research process is to navigate the intersection of ML and ethics (Akter et al., 2022). In growth marketing or any other ML application, data bias occurs when the available data does not consider the variables that adequately capture the overall prediction phenomenon. The presence of data bias impacts the algorithms' overall performance as the model training is compromised. This leads to inaccurate predictions as the data does not correctly reflect the predictive analysis but instead works on a set of biased datasets (van Giffen et al., 2022). Though data bias is a significant problem, it remains unresolved. This is because the datasets used in ML and AI applications are large and complex, always including outliers and external factors. These external factors are such that they do not truly depict the actual predictive scenario. Therefore, it is essential to stay aware of the degree of bias a dataset may have. This can help recognize the data biases before making data models production-ready, thereby handling costs and efforts (van Giffen et al., 2022).

Algorithmic bias in the marketing domain can have serious consequences, such as violating the right to privacy, leading to discriminatory decision-making, and limiting opportunities for certain groups of people (Grewal, Hulland, Kopalle, & Karahanna, 2020). For example, many websites use the news feed algorithm and gather information about user preferences and other sensitive attributes such as race, gender, sexual orientation, and opinions (Fosch-Villaronga et al., 2021). This information is then used to improve online behavioral advertising and enhance customer experience. It can sometimes enforce gender stereotypes in a marketing strategy that perpetuates the idea that men and women are different kinds of people. For example, if an algorithm is trained on a historical gender dataset that shows women are more likely to be interested in certain products, it may show those products more often to women than to men, even if both genders are equally likely to be interested. The online advertising market also has a poor track record in racial diversity. Azer, Anker, Taheri, and Tinsley (2023) discussed race and ethnicity discrimination as a major threat to the public interest because it directly violates the Civil Rights Act of 1964, which protects all customers against racially discriminatory advertising. An algorithm should not be trained on a dataset that shows that certain races are more likely to buy certain products or respond to certain online advertisements. If the algorithm is not designed to account for this bias, it could end up showing those products or advertisements more often to people of that race.

More broadly, whenever a feature or set of variables has a detrimental impact on the overall learning of ML models and produces inaccurate recommendations or predictions (Lambrecht and Tucker, 2019), it may lead to failed marketing campaigns. If ML algorithms are biased toward a few features or variables, they may not be able to capture the full range of customer behavior or preferences. This can result in incomplete or inaccurate insights that can hinder effective marketing decisions (Akter et al., 2022). It is important for marketers to be aware of these biases and work to minimize them in the ML model development process, as outlined by van Giffen et al. (2022). Working with AI systems is an investment. To ensure that the AI systems set up in an organization do not produce biased or inaccurate results, marketers must ensure that the AI model results are evaluated by prioritizing the business goals. This will help marketers optimize returns and reduce data integrity risk.

4.5. Causal machine learning

Most of the ML research in marketing relies on predictive ML to identify patterns within data for predicting a specific outcome (e.g., sales or customer engagement). To do so, predictive ML algorithms use a portion of the data to train models that predict the outcome based on identified patterns. Another portion of the data is then used to determine the model with the best performance, where performance is evaluated based on the accuracy of predicted outcomes compared to actual outcomes (e.g., Vermeer et al., 2019). To maximize the predictive power of a model and minimize prediction error, predictive ML algorithms make a trade-off between bias and variance (Belkin, Hsu, Ma, & Mandal, 2019). Bias refers to the systematic deviation of the chosen model from the true predictive model, and variance refers to the sensitivity of the predictions to which data is used for training the model. By striking the right balance between bias and variance, predictive ML algorithms can improve the accuracy of their predictions (for a detailed discussion, see Langen & Huber, 2023).

Overall, a large body of marketing research demonstrates the effectiveness of predictive ML algorithms in identifying patterns within complex datasets and predicting outcomes. However, this approach has some notable limitations. For example, when multiple variables capture the same relevant predictive feature, ML algorithms may identify some of these variables as relevant predictors while ignoring others. This can happen even if the omitted variables have a causal impact on the outcome of interest. For example, variables that do not directly or only modestly affect the outcome may enter the predictive model as relevant predictors simply because they are correlated with other variables that do affect the outcome. As a result, these other variables may play little or no role in the predictive model, even though they have a causal impact on the outcome (Langen & Huber, 2023). This makes predictive ML

[3] LIME fits a simple model around a prediction to create a local explanation while SHAP uses game theory to measure the importance of each feature. Both methods are used to understand how features affect a predicted outcome, though each method uses a slightly different procedure.

models unsuitable for answering questions about an intervention's effects on an outcome. Answering such questions requires a causal analysis, which considers the underlying mechanisms by which variables affect outcomes. Predictive models are not designed to provide this type of information, so they may not be the best tool for decision support in situations where causal analysis is needed, such as when designing a marketing campaign. Given the predictive focus, little attention has been paid to endogeneity concerns when developing ML methods. Issues such as selection, omitted variables, and simultaneity are also typically ignored in ML (Ma & Sun, 2020). This lack of causality makes using ML for marketing decisions and actions challenging.

After the initial surge of enthusiasm surrounding the novel opportunities presented by predictive ML for researchers and businesses, commentators increasingly stress the importance of integrating causal inference and ML (e.g., Hair & Sarstedt, 2021). In response to the shortcomings of predictive ML, a rapidly growing body of research in econometrics and statistics (e.g., Wager & Athey, 2018) has been focusing on causal ML algorithms. By considering causal relationships between variables, causal ML algorithms can provide more robust and reliable insights into the underlying mechanisms of patterns found in data. This contrasts with traditional predictive ML, which focuses solely on prediction without regard for causality. Therefore, it is not surprising that the use of causal ML in marketing research is rapidly gaining popularity. Although still in its early stages, the literature utilizing causal ML approaches is growing steadily. For instance, Zhang and Luo (2023) used cluster-robust causal forests in their study to model restaurant survival based on social media photos. Langen and Huber (2023) used causal forests to examine the causal effects of receiving coupons on customer spending. Then they employed optimal policy learning to determine which customer segments should be targeted with coupons. Guo, Sriram, and Manchanda (2021) explored the impact of information disclosure on industry payments to physicians and used causal forests to evaluate the expected individual-level effect heterogeneity. Lastly, Narang, Shankar, and Narayanan (2019) applied causal forests to analyze the heterogeneity across shoppers regarding the impact of mobile app failures on consumers.

While these developments are promising and necessary, recent work by Gordon et al. (2023) strikes a more cautionary tone regarding the workings and limits of causal ML. They analyzed 663 experiments on Facebook to investigate whether non-experimental approaches can accurately recover the causal effects of advertising observed in the experiments. The study used two non-experimental methods, double/debiased ML and stratified propensity score matching, to recover the experimental effects observed in randomized controlled trials. However, neither method performed well, indicating significant relative measurement errors. The authors suggest that this is a data rather than a model problem. This shows that the quality and nature of the data fed into causal ML plays a crucial role in its effectiveness. Therefore, causal ML is not a magic solution and often falls short of the gold standard in causal inference: randomized controlled trials. Instead, it should be understood as an additional tool in marketing researchers' toolbox that, combined with other methods, can tackle novel marketing problems.

Because causal ML research in marketing is still in its early stages, there are many areas that offer potential for future exploration. For example, we anticipate that some of the seminal findings achieved through predictive ML will be subject to closer examination using causal ML techniques. Additionally, it is important to understand how causal ML can be leveraged to improve its accuracy and bring it closer to the standards set by randomized controlled trials. Therefore, careful consideration of the best practices for using causal ML in marketing would be advantageous for advancing the field.

5. Conclusion and acknowledgments

We hope this overview article and the papers in this Special Issue serve as an impetus for further research on the important topic of ML in marketing. In addition to summarizing the insights from the published papers, we also discussed important future research directions related to the themes of (1) automated machine learning, (2) data privacy and security, (3) model interpretability, (4) algorithm fairness, and (5) causal machine learning. Selected research questions for ML in marketing derived from these discussions are summarized in Table 2. In addition, broader ML themes might also benefit from a marketing perspective. For example, how can we ensure that ML models are producing ethical and socially responsible models and predictions (De Cremer, 2020; De Cremer & Kasparov, 2022)? Deep learning seems to change how prediction models are developed (Kaur & Sharma, 2023) – what will these techniques' impact on marketing analytics be in the near future? And, importantly, how will ChatGPT and similar advanced AI technologies change marketing research (van Dis, Bollen, Zuidema, van Rooij, & Bockting, 2023)?

We want to thank all contributors to this Special Issue. We received a large number of competitive submissions, of which finally 16 submissions were published. We want to note that when there was any potential conflict of interest between the authors of articles due to, e.g., co-authorship, a co-editor of the Special Issue handled the article without any conflicts of interest. We are grateful to all reviewers who assisted us in curating this Special Issue. Their authoritative adjudication of the manuscripts across several rounds of reviews provided extensive and valuable suggestions. Finally, we thank the former co-editors of the *Journal of Business Research*, Naveen Donthu, and Anders Gustafsson, for providing us the opportunity to compile this Special Issue and for their continuous support throughout the review process, and the current editor Mirella Kleijnen for providing extensive feedback on our editorial.

Last but not least, we warmly thank all the authors who submitted their manuscripts for consideration for inclusion in the *Journal of Business Research*. We appreciate and are grateful for the authors' desire to share their knowledge and experience with the Journal's readers—and for having their views put forward to be challenged by the reviewers and their peers. We are confident the articles in this Special Issue contribute to our understanding of ML in marketing and anticipate that you will concur when you read them.

CRediT authorship contribution statement

Dennis Herhausen: Conceptualization, Writing – original draft. **Stefan F. Bernritter:** Conceptualization, Writing – original draft. **Eric W.T. Ngai:** Conceptualization, Writing – original draft. **Ajay Kumar:** Conceptualization, Writing – original draft. **Dursun Delen:** Conceptualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Boston, Massachusetts: Harvard Business Review Press.

Airani, R., & Karande, K. (2022). How social media effects shape sentiments along the twitter journey? A Bayesian network approach. *Journal of Business Research, 142*, 988–997.

Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research, 144*, 201–216.

Azer, J., Anker, T., Taheri, B., & Tinsley, R. (2023). Consumer-Driven racial stigmatization: The moderating role of race in online consumer-to-consumer reviews. *Journal of Business Research, 157*, Article 113567.

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences, 116*(32), 15849–15854.

Boom, B. H., & Bitner, M. J. (1981). *Marketing of Services*. Chicago: America Association.

Bringer, J., Chabanne, H., & Patey, A. (2013). Privacy-preserving biometric identification using secure multiparty computation – An overview and recent trends. *IEEE Signal Processing Magazine*, 42–52.
Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science, 24*(4), 595–615.
Cunha, M., Mendes, R., & Vilela, J. (2021). A survey of privacy-preserving mechanisms for heterogeneous data types. *Computer Science Review* (Vol. 31), Article 100403.
Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48* (1), 24–42.
De Cremer, D. (2020). What does building a fair AI really entail? *Harvard Business Review*.
De Cremer, D., & Kasparov, G. (2022). The ethical AI-paradox: Why better technology needs more and not less human responsibility. *AI and Ethics, 2*(1), 1–4.
De Luca, L. M., Herhausen, D., Troilo, G., & Rossi, A. (2021). How and when do big data investments pay off? The role of marketing affordances and service innovation. *Journal of the Academy of Marketing Science, 49*(4), 790–810.
Delen, D. (2020). *Predictive analytics: Data mining, machine learning and data science for practitioners*. Upper Saddle River, NJ, USA: FT Press.
Esmeli, R., Bader-El-Den, M., & Abdullahi, H. (2022). An analysis of the effect of using contextual and loyalty features on early purchase prediction of shoppers in e-commerce domain. *Journal of Business Research, 147*, 420–434.
Feng, Y., Yin, Y., Wang, D., & Dhamotharan, L. (2022). A dynamic ensemble selection method for bank telemarketing sales prediction. *Journal of Business Research, 139*, 368–382.
Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021). A little bird told me your gender: Gender inferences in social media. *Information Processing & Management, 58*(3), Article 102541.
Ghouri, A. M., Mani, V., ul Haq, M. A., & Kamble, S. S. (2022). The micro foundations of social media use: Artificial intelligence integrated routine model. *Journal of Business Research, 144*, 80–92.
Gordon, B. R., Moakler, R., & Zettelmeyer, F. (2023). Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. *Marketing Science, 42*(4), 768–793.
Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence, 4*, 752558.
Grewal, D., Hulland, J., Kopalle, P. K., & Karahanna, E. (2020). The future of technology and marketing: A multidisciplinary perspective. *Journal of the Academy of Marketing Science, 48*, 1–8.
Guo, T., Sriram, S., & Manchanda, P. (2021). The effect of information disclosure on industry payments to physicians. *Journal of Marketing Research, 58*(1), 115–140.
Hagen, L., Uetake, K., Yang, N., Bollinger, B., Chaney, A. J. B., Dzyabura, D., & Sudhir, K. (2020). How can machine learning aid behavioral marketing research? *Marketing Letters, 31*(4), 361–370.
Hair, J. F., Jr, & Sarstedt, M. (2021). Data, measurement, and causal inferences in machine learning: Opportunities and challenges for marketing. *Journal of Marketing Theory and Practice, 29*(1), 65–77.
Herhausen, D., Miočević, D., Morgan, R. E., & Kleijnen, M. H. (2020). The digital marketing capabilities gap. *Industrial Marketing Management, 90*, 276–290.
Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50.
Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data, 10*(1), 1–23.
Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review, 2*(1), 3–10.
Kolomoyets, Y., & Dickinger, A. (2023). Understanding value perceptions and propositions: A machine learning approach. *Journal of Business Research, 154*, Article 113355.
Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management, 90*, 493–507.
Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science, 65*(7), 2966–2981.
Langen, H., & Huber, M. (2023). How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign. *PLoS One1, 18*(1), Article e0278937.
Latinovic, Z., & Chatterjee, S. C. (2022). Achieving the promise of AI and ML in delivering economic and relational customer value in B2B. *Journal of Business Research, 144*, 966–974.
Liu, H., Tan, K. H., & Pawar, K. (2022). Predicting viewer gifting behavior in sports live streaming platforms: The impact of viewer perception and satisfaction. *Journal of Business Research, 144*, 599–613.
Liu, L., Dzyabura, D., & Mizik, N. (2020). Visual listening in: Extracting brand image portrayed on social media. *Marketing Science, 39*(4), 669–686.
Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research, 56*(6), 918–943.
Luo, X. M., Tong, S. L., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*(6), 937–947.
Lutz, B., Pröllochs, N., & Neumann, D. (2022). Are longer reviews always more helpful? Disentangling the interplay between review length and line of argumentation. *Journal of Business Research, 144*, 888–901.
Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing, 37*(3), 481–504.
Mehmood, A., Natgunanathan, L., Xiang, Y., Hua, G., & Gu, S. (2016). Protection of big data privacy. *IEEE Access, 4*, 1821–1834.
Mitchell, T. M. (1997). *Machine learning*. Burr Ridge, IL: McGraw Hill.
Narang, U., Shankar, V., & Narayanan, S. (2019). *The impact of mobile app failures on purchases in online and offline channels*. Working Paper. Retrieved, April 5, 2023, from: https://www-2.rotman.utoronto.ca/userfiles/seminars/marketing/files/4_Narang_Unnati_JobMarketPaper_2019(1).pdf.
Ngai, E. W., & Wu, Y. (2022). Machine learning in marketing: A literature review, conceptual framework, and research agenda. *Journal of Business Research, 145*, 35–48.
Ordenes, F. V., & Silipo, R. (2021). Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications. *Journal of Business Research, 137*, 393–410.
Park, U., & Jain, A. K. (2010). Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security, 5*(3), 406–415.
Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–688.
Philp, M., Jacobson, J., & Pancer, E. (2022). Predicting social media engagement with computer vision: An examination of food marketing on Instagram. *Journal of Business Research, 149*, 736–747.
Potrawa, T., & Tetereva, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research, 144*, 50–65.
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144).
Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing, 37*(1), 15–26.
Sengupta, P., Biswas, B., Kumar, A., Shankar, R., & Gupta, S. (2021). Examining the predictors of successful Airbnb bookings with Hurdle models: Evidence from Europe, Australia, USA and Asia-Pacific cities. *Journal of Business Research, 137*, 538–554.
Sheth, J., & Kellstadt, C. H. (2021). Next frontiers of research in data driven marketing: Will techniques keep up with data tsunami? *Journal of Business Research, 125*, 780–784.
Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479.
Vaccaro, L., Sansonetti, G., & Micarelli, A. (2021). An empirical review of automated machine learning. *Computers, 10*(1), 11.
van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226.
van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research, 144*, 93–106.
Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing, 36*(3), 492–508.
Volkmar, G., Fischer, P. M., & Reinecke, S. (2022). Artificial Intelligence and Machine Learning: Exploring drivers, barriers, and future developments in marketing management. *Journal of Business Research, 149*, 599–614.
Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242.
Wang, T., He, C., Jin, F., & Hu, Y. J. (2022). Evaluating the effectiveness of marketing campaigns for malls using a novel interpretable machine learning model. *Information Systems Research, 33*(2), 659–677.
Yoganarasimhan, H. (2020). Search personalization using machine learning. *Management Science, 66*(3), 1045–1070.
Zhang, M., & Luo, L. (2023). Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp. *Management Science, 69*(1), 25–50.
Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. S. (2022). More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering, 34*(6), 2824–2843.