Augmenting the Future: An Exploratory Analysis of the Main Resources, Use Cases and Implications of Augmented Analytics

Thuy Duong Oesterreich  
*Osnabrück University, toesterreich@uni-osnabrueck.de*

Eduard Anton  
*University of Osnabrück, eduard.anton@uos.de*

Feipeng Xu  
*Osnabrück University, feixu@uni-osnabrueck.de*

Follow this and additional works at: [https://aisel.aisnet.org/ecis2021_rp](https://aisel.aisnet.org/ecis2021_rp)

**Recommended Citation**  
[https://aisel.aisnet.org/ecis2021_rp/19](https://aisel.aisnet.org/ecis2021_rp/19)
AUGMENTING THE FUTURE: AN EXPLORATORY ANALYSIS OF THE MAIN RESOURCES, USE CASES, AND IMPLICATIONS OF AUGMENTED ANALYTICS

Research Paper

Oesterreich, Thuy Duong, Osnabrück University, Osnabrück, Germany, thuyduong.oesterreich@uni-osnabrueck.de
Anton, Eduard, Osnabrück University, Osnabrück, Germany, eduard.anton@uni-osnabrueck.de
Xu, Feipeng, Osnabrück University, Osnabrück, Germany, feixu@uni-osnabrueck.de

Abstract

Recently, augmented analytics has increasingly gained attention as one of the more advanced, novel approaches for handling big data. Based on machine learning and natural language processing, augmented analytics benefits from recent advancements in the artificial intelligence field to automate the analytics cycle. Despite the various benefits that augmented analytics offers for business and society, research on this topic is scarce to date. Based on the IT business value model, we examine the role of technological and social resources as well as the main use cases of augmented analytics. Therefore, we combine quantitative text mining with qualitative content analysis for an exploratory study of 350 academic and practical publications as well as 49 datasets of companies offering augmented analytics software and services. The findings contribute to the body of knowledge by enhancing the understanding of the augmented analytics concept, uncovering prevalent research gaps, and highlighting future research directions.

Keywords: Augmented analytics, Artificial intelligence, Big data analytics, IT business value model, Text mining.

1 Introduction

The increasing digitization is leading to huge, ever-growing data volumes (Grover et al., 2018). In research and practice, the term big data is used to describe these tremendous very large and complex datasets from various sources (Chen et al., 2012). Manual approaches to explore and detect patterns, make sense of data, and draw conclusions are becoming increasingly challenging and prone to bias due to this extensive data growth (Sallam et al., 2017). Instead, advanced techniques for the storage, management, analysis, and visualization of data are required, so-called big data or business analytics (BDA) (Chen et al., 2012). Among the common concepts of BDA, augmented analytics has recently gained the attention of researchers and practitioners as one of the more advanced, novel approaches for handling big data (Davenport & Harris, 2017; Prat, 2019). Augmented analytics was first forecasted as one of the emerging technology trends in 2017 under the label “augmented data discovery” in “Gartner’s Hype Cycle for Emerging Technologies” (Panetta, 2017). In the “Top 10 Strategic Technology Trends for 2019,” it even ranked second (Panetta, 2018). Augmented analytics is generally understood as BDA enabled by artificial intelligence (AI). More specifically, augmented analytics automates the analytics cycle by applying machine learning (ML) and natural language processing (NLP) to benefit from recent advancements in the AI field (Prat, 2019).
Despite various benefits for business and society, augmented analytics is still in its early stages, with many open questions regarding its socio-technical system, i.e., technologies, people, processes, and their interactions (Prat, 2019). By examining technological and social issues within the analytics cycle as well as its main application fields in business and society, we intend to enhance the understanding of this concept and its implications for research and practice. In line with the common view in the BDA literature, we consider augmented analytics as a socio-technical system consisting of technological components such as software and services and social components, including human resources as well as management capabilities (Grover et al., 2018). By viewing augmented analytics as a socio-technical concept through the lens of the IT business value model (Melville et al., 2004; Schryen, 2013) and the BDA value creation model (Grover et al., 2018), we aim to answer the following research questions (RQs) in an exploratory manner:

**RQ1:** What are the main conceptual characteristics of augmented analytics, and to what extent does this concept contribute to automating the analytics cycle?

**RQ2:** To what extent does the role of technological and human resources as well as management capabilities change in the augmented analytics concept?

**RQ3:** What are the main fields of application and use cases of augmented analytics, and what implications does the emergence of augmented analytics entail for business and society?

By addressing these RQs, we answer the call for more research to understand and interpret the recent emergence of BDA in the IS field (Baesens et al., 2016). The remainder of this paper is structured as follows. In Section 2, we outline the theoretical background and highlight the recent emergence of augmented analytics as one advanced concept of BDA. In Section 3, we describe the research methods that help us to answer the RQs. Based on a sample of 350 academic and practical publications as well as 49 datasets of companies that offer augmented analytics software and services, we combine a quantitative text mining approach in R with a qualitative content analysis for an exploratory analysis of the main resources, use cases, and implications of augmented analytics. The results are presented in Section 5, followed by a detailed discussion in Section 6. Concluding remarks are offered in Section 7.

## 2 Big Data and the Emergence of Augmented Analytics

In recent decades, terms such as BDA have become popular, serving to describe the extensive growth of data of various types and sources that require advanced storage, management, analysis, and visualization techniques (Chen et al., 2012). BDA is even praised as “the most significant 'tech' disruption in business and academic ecosystems since the meteoric rise of the Internet and the digital economy” (Agarwal and Dhar, 2014, p. 443). By detecting new patterns and correlations, the use of BDA enables firms to gain valuable insights and draw conclusions from big data (Walker, 2014), which in turn allows for better predictions and more informed decisions (McAfee et al., 2012). With conventional BDA techniques, it is not possible to explore all existing patterns when data volume and the number of different variables are large. Rather, even discovering the most relevant and actionable results is time-consuming and difficult (Kronz, 2019). In recent years, advances in the field of AI have enabled the application of BDA through higher levels of automation. The most common BDA concepts include techniques with differing levels of intelligence, such as descriptive, predictive, prescriptive, and autonomous analytics. The different levels of intelligence, in turn, are expected to contribute to different levels of competitive advantage (Davenport & Harris, 2017). The highest level of intelligence is achieved by autonomous analytics or augmented analytics techniques that employ AI to create self-learning and self-optimizing models with less involvement from human analysts (Davenport and Harris, 2017).

Compared to conventional BDA, augmented analytics offers firms various advantages in many respects. Aside from faster and better decisions (LaPlante, 2019, p. 9), augmented analytics is considered a remedy for the skills shortage in the data science market. At the same time, the potential of augmented analytics to automate a great portion of analysts’ tasks constitutes a challenge to the aforementioned view of BDA as a socio-technical system. Despite the various benefits, challenges,
and further implications that augmented analytics might entail for business and society, the focus of prior research is on technological artifacts (Gole and Shiralkar, 2020; Razzak Malik et al., 2017) or conceptual descriptions of augmented analytics (Andriole, 2019; LaPlante, 2019; Prat, 2019). Notwithstanding the urgent need for more research on issues related to technologies, people, processes, and their interactions concerning augmented analytics (Prat, 2019), the short list of related works indicates that research in this field is limited to date. We aim to close this gap by examining the role of technological and social resources within the analytics cycle as well as the main use cases in which augmented analytics can be of value for business and society.

3 Research Method

We employ a multi-step research approach that enables us to study the topic of interest in an exploratory manner. To answer the RQs, we aim to examine the patterns in academic and practical publications as well as the market of augmented analytics. The first step of our data collection process consisted of a systematic literature search. Therefore, we followed the recommendations of Webster and Watson (2002, p. 16) to search for literature in a manner that allows us to “accumulate a relatively complete census of relevant literature.” Similarly, vom Brocke et al. (2009) underline the importance of a comprehensive literature search to capture all relevant publications.

Given that big data analytics is considered an interdisciplinary topic that can be located at the interfaces between IS research and related fields of computer science, marketing, management, communication, and mathematics (Chen et al., 2012), we conducted a rather broad literature search in the interdisciplinary databases EBSCOhost, ScienceDirect, and SpringerLink. We used the search term “augmented analytics” and other common synonyms such as “autonomous analytics”, “AI-driven analytics”, and “cognitive analytics” that are often used to describe the AI-driven nature of augmented analytics. To capture as many publications as possible, we did not restrict the database search to titles or abstracts, but searched in the articles’ full text.

The literature search resulted in 803 articles that we checked for relevance by reading titles and abstracts. When the relevance of an article could not be assessed in this way, we referred to the full text. In order to avoid bias, we consistently applied a set of predefined inclusion and exclusion criteria to avoid selection biases. We only included publications focusing on augmented analytics or addressing the impact of AI on BDA. Additionally, we included case studies and white papers that present interesting findings on use cases and vendors of specific augmented analytics tools and services. Publications were excluded when they only briefly mention augmented analytics as a buzzword or present fee-based market forecasts without further insights into the topic. With respect to publication type and language, we only included publications that are written in English and published in an academic or practical outlet, including journal and conference papers, working papers, technical reports, and white papers. Two researchers conducted this assessment step independently yielding a substantial inter-rater reliability (Cohen’s Kappa) of 0.79, thus a satisfying level of agreement (Landis and Koch, 1977; LeBreton and Senter, 2008). Based on this assessment step, we excluded 453 publications not meeting our inclusion criteria and thus being irrelevant to our study. For the subsequent content analysis, we downloaded the full texts of the remaining relevant 350 academic and practical publications. Furthermore, we searched for vendors on Crunchbase, an internet database provider for business information (Crunchbase Inc., 2020) to gain a holistic and valid overview of the market of augmented analytics products and services. For this purpose, we used Crunchbase’s query function and again applied the abovementioned search terms. Furthermore, we limited the results to actively operating companies for a timelier picture and enriched the data with information from the company websites.

We used quantitative content analysis to draw “replicable and valid inferences” (Krippendorff, 1989, p. 403) from the corpus of 350 relevant studies and the company information extracted from Crunchbase. The applied quantitative content analysis techniques include inductive text analytical procedures to identify recurring text patterns in the literature sample (Roberts, 1997). Text analytical procedures are frequently used in the IS literature to derive thematic (Anton et al., 2020; Sidorova et
al., 2008; Weigel et al., 2013) and semantic conclusions (Yim et al., 2020; Zhang et al., 2016) over a large body of text data. We applied several text analytical procedures, namely a) frequency analysis for extracting relevant features from the text corpus, b) keyword-in-context and network analyses to gain a deeper insight and semantic understanding of the extracted features, and c) named entity recognition to identify and classify named entities (i.e., organizations that use augmented analytics or are associated with the technology). Therefore, we used the R programming language and dedicated text mining and visualization packages (e.g., quanteda, spaceyr, RTextTools, and ggplot) to gain insights into how and when companies can achieve business value through augmented analytics (use cases). To ensure transparency and reliability in our quantitative approach, we provide supplementary information at https://bit.ly/3rnIPcx. The analysis simultaneously focused on contextual factors (company, industry, and country) and on IS and non-IS investments required to generate business value (Schryen, 2013). Prior to the analysis, we conducted several pre-processing steps to reduce noise and dimensionality from the data by removing punctuation, numbers, and English stop words (e.g., “a”, “the”). The results from the quantitative content analysis serve as the basis for the subsequent concept-centered qualitative analysis of the literature (Webster and Watson, 2002). Two researchers independently performed deductive analyses of the full texts guided by the concept matrix. They coded the texts to the concepts in accordance to the guidelines on the systematic, rule-based qualitative text analysis approach of Mayring (2000). This involved screening the full texts for the most frequent unigrams and bigrams that form the concept matrix (e.g., technological, human, organizational resources, application fields, and use cases). We validated the consistency of the coding process using Cohen’s Kappa which again revealed a substantial agreement of 0.72 (Landis and Koch, 1977).

4 IT Business Value of Augmented Analytics

As stated in previous sections, we aim to examine the role of technological and social resources of augmented analytics within the analytics cycle as well as its main use cases and implications. Therefore, we consider augmented analytics as a socio-technical concept through the lens of the IT business value model (Melville et al., 2004; Schryen, 2013) as well as the BDA value creation model (Grover et al., 2018). This view implies that not only technological components such as products and services, but also social elements such as human resources and management capabilities are relevant for creating value from augmented analytics (Grover et al., 2018). The socio-technical nature of BDA in general and augmented analytics in particular can be confirmed when analyzing the results of the network analysis based on the most frequent unigrams and bigrams (two-word-phrases) extracted from the text corpus of the academic and practical publications in Figure 1.

![Figure 1. Relationships between the most frequent unigrams and bi-grams extracted from the text corpus (left: unigrams; right: bigrams).](image_url)
foundations of augmented analytics, such as machine, learning, machine learning, intelligence, artificial intelligence, natural, language, natural language, neural networks, and cognitive computing. The results also reveal that data play a major role in the augmented analytics concept, with terms such as data, big data, structured data, and unstructured data frequently occurring within the text corpus. A process-oriented view of augmented analytics is also indicated by terms that describe the different stages of the analytics cycle, such as process, discovery, data discovery, preparation, data preparation, analysis, visualization, and data visualization. The frequency of unigrams and bigrams such as resources, capabilities, analytics capabilities, and data scientists also clearly underlines the major role that social elements, i.e., human resources and management capabilities, play within the discourse on the concept of augmented analytics. Terms such as automated, insights, and actionable insights are frequently used, which emphasizes the value of augmented analytics. Interestingly, the terms smart cities, healthcare, and supply chain are among the most frequent terms, again indicating the relevance of augmented analytics to these application fields. We delve deeper into the main use cases in these fields in Section 5.

Another interesting view is offered when investigating the relationships between the extracted terms within the network analysis in Figure 1. The network analysis enables us to visually study the relationships between the most common terms and the term augmented analytics in a more effective manner. The frequency of the co-occurrence is indicated by the thickness of the arcs. For example, the thick line between the bigrams big data and machine learning indicates that these two terms are most often used together in the publications, followed by the combinations of the bigrams artificial intelligence and machine learning, as well as machine learning and cognitive computing. The term augmented analytics is often associated with terms such as big data, machine learning, natural language, data preparation, and data scientists in the same document. However, although not directly connected to some other frequently used terms in the network, augmented analytics can still be associated with terms such as cognitive computing or the internet of things via other common terms like big data, artificial intelligence, and machine learning. In the next sections, we delve deeper into the questions of how technological advances in the AI field have indeed challenged the socio-technical nature of the BDA.

4.1 Technological resources enabling augmented analytics

Melville et al. (2004) categorize technological IT resources into IT infrastructure and business applications that utilize the infrastructure. This classification also applies to augmented analytics. Our analysis of the literature uncovers two broad areas that form the technological fundament for effectively exploiting the benefits of augmented analytics, namely (a) the data operations that refer to the infrastructure to integrate and manage data, and (b) the algorithms and software to conduct the pre-processing and analytical procedures utilizing the data. The operational foundation for the use of big data, regardless of the analytical maturity level, is the availability of data as well as its preparation and management throughout the data lifecycle to enable predictive and causal inferences about the business and contribute to its development (Baesens et al., 2016). Conventional business intelligence solutions rely on data warehouses based on structured data to store and extract the required information. In such solutions, data curation to integrate data from business applications and other sources is a manual and time-consuming process, as the data schema is pre-specified, meaning that deviations in the data structure have to be adjusted a priori. However, the structured nature facilitates the subsequent extraction and analysis steps (Prat, 2019; Pribisalić et al., 2019). Well-known analytics models such as the CRISP-DM (Shearer, 2000) or the analytics cycle based on Prat (2019) establish data processing on the basis of a business problem or opportunity. Consequently, efforts in data preparation, data analysis, or model deployment are only called for when the data are actually needed so that decision makers can act on the results of the analysis. Therefore, to reduce unnecessary data curation steps and human-dependent intervention, merging and streaming different data types from various sources more flexibly and with less preliminary effort must be possible. One such possibility is provided by data lakes that collect relational business information and unstructured data from IoT or social media without being bound to a data schema, but rather store them in their existing structure.
according to the data plan of their distributed sources (schema on read) (Shepherd et al., 2018). Although data lakes reduce the effort and costs of the preliminary data preparation process, they do not eliminate it; instead, they collect all types of data in their raw form and make them available for further processing on demand (Shepherd et al., 2018). However, collecting data from different sources and streaming and processing data for augmented analytics require an infrastructure of distributed systems and the know-how for distributed computing (Munappy et al., 2020).

Typically, a team of business developers, data scientists, and software engineers is required to access and extract the required information from data, especially if it is unstructured (Anton et al., 2020). However, augmented analytics specifically streamlines these processing phases that lie between the business problem and the decision to respond to it (Prat, 2019). The increasing automation of these pre-processing and analytical steps is primarily achieved through the use of algorithms and software associated with AI. Artificial intelligence is an umbrella term for technologies and techniques ascribing to software abilities that can be interpreted as intelligent (Russell and Norvig, 2016). The key techniques operationalizing AI in augmented analytics are ML and NLP. Machine learning comprises a set of different clustering, classification, and regression algorithms that are used for training models that are applied for predictions on data with similar data features as the training set (Russell and Norvig, 2016; Wang et al., 2019). Therefore, the strengths of ML lie in the analysis of structured data. Natural language processing subsumes algorithms to explore, “understand and manipulate natural language text or speech” (Liu et al., 2017, p. 1), and thus possesses abilities for analyzing unstructured data. Artificial intelligence assists in creating an automated pipeline of preprocessing and modeling tasks harnessing the power of augmented analytics. The uniqueness of augmented analytics is that it transforms the necessary phases within the analytics cycle that were previously mainly in the hands of experts. As the technology matures, it also addresses areas of early process phases, where business problems and opportunities are identified (Prat, 2019), and late phases, where decision makers typically react according to the results of the analyses (Baesens et al., 2016). Thus, AI eventually enables self-service functions for business users with no or marginal data science expertise (Pribisalić et al., 2019).

4.2 Market analysis: Augmented analytics solutions

Prat (2019) provides an overview of AI applications that enable automation within the analytics cycle, spanning data profiling and transformation to discovery and modeling tasks. We augment the market overview of Prat (2019) by identifying 49 companies that offer AI-based automated analytical solutions. For this purpose, we used the named entity recognition method and searched for organizations on Crunchbase (cf. Figure 2).

These service or software providers include augmented analytics in their portfolio (on-premise or cloud solutions) and mainly operate in the USA (63%), followed by the UK (8%) and Israel (6%). Most of the organizations have less than 50 employees (47%, e.g., Beacon Intelligence and Topos Labs), but the sample also includes major players with more than 1,000 employees (24%, e.g., SAP, IBM, and Salesforce). Few of these organizations concentrate on specific industries such as healthcare (12%, e.g., MaxQ-AI and Data2Life), finance (8%, e.g., Opal and Finactum), or telecommunications (4%, e.g., Fleet Space Technologies), but rather offer analytics software and services for industry-spanning use cases (63%, e.g., Sisense and MicroStrategy). However, the maturity and scope of the solutions vary among the providers, with software covering only particular phases of the analytics cycle or constituting comprehensive autonomous platform solutions for the analytical process. Our list of functionalities and solutions along the analytics cycle is only an overview of the augmented analytics market and does not claim to be exhaustive.

Data preparation is one of the most complex and tedious tasks that lays the foundation for further data analysis and model deployment (Prat, 2019; Pribisalić et al., 2019). Often cleansing and analytical tasks have to be programmed and executed manually step by step, as for example in our quantitative analysis using the programming language R. Solutions such as Decooda’s AI-based engine can automate these steps within a pipeline of data organization, noise filtering, transforming, and enriching
the information with content from the unstructured and structured data of other sources (Decooda, 2020). As effective as the capabilities may be, the data preparation has to be tailored to the business problem. Navigation around this usually requires a human pilot (Prat, 2019). Tools such as IBM’s InfoSphere Advanced Data Preparation takes this into account by providing ML-based data preparation recommendations via a dashboard that steers the user through the process (IBM, 2020).

Figure 2. Providers of augmented analytics solutions.

Providers such as Qlik Technologies and Tableau offer AI-based data discovery platforms, including tools for analyzing and visualizing data. The platform solutions provide the capacity to instantly create dashboards with recommended charts and illustrations based on the input data, revealing patterns and relationships between data features (Qlik, 2020; Tableau, 2020). Such discovery tools can be used for feature extraction, thus model development. Other augmented analytics platforms such as DataRobot (DataRobot, 2020) and H2O.ai (H2O.ai, 2020) even enable feature engineering by extracting or transforming information from existing features. Pre-specified models can be automatically selected based on model libraries that correspond to the feature composition of a specific use case, and they can be subsequently tested and validated based on specific metrics (Anodot, 2020). Furthermore, this has implications for model deployment; with predefined and autonomous trained models, even employees without data science knowledge can import models into the production environment (to avoid interdependencies with other models or data, this should be applied in virtualized containers e.g., via Docker (2020)) (DataRobot, 2020). In addition, monitoring tools can control the environment. Anodot uses ML to autonomously monitor the entire process of ML development, deployment, and operations to control for anomalies or bad releases that affect performance (Anodot, 2020).

Software solutions can already cover most phases in the analytics cycle and reduce human involvement, but solutions for the phases of decision, action as well as business problem or opportunity identification are more difficult to automate (Prat, 2019). Many tools support decision making or problem identification through data insights, such as SAP’s Smart Discovery, that can process user queries through NLP to access and analyze data and respond to a query via data
visualization (Ivain, 2018). Yet, our market analysis has shown that the technological maturity level in the decision-making process currently only allows for providing recommendations for decisions and problem solving. The technological maturity level is inadequate for outsourcing responsibility and data governance to algorithms.

4.3 Human and organizational resources

Aside from technological resources, human and organizational resources play a major role in the socio-technical concept of BDA. Grover et al. (2018) propose that BDA infrastructure such as assets, analytics portfolio, and human talent are the main resources that must be combined with the organizational capabilities to create value from BDA. In a similar manner, Akter et al. (2016) consider BDA technology capability, BDA talent capability, and BDA management capability as the three principal resources and capabilities necessary to generate value. Previous research shows that human and organizational factors play a significant role in creating value from BDA and that competitive advantage cannot be achieved by simply adopting a BDA tool (Gupta & George, 2016; Someh et al., 2019). Rather, firms need to combine their resources and develop unique capabilities (Božič & Dimovski, 2019a; Ferraris et al., 2019). Some studies state that social aspects are even more essential for enhancing firm performance than technological ones (Božič & Dimovski, 2019b). The reason is that BDA tools constitute non-core resources that can be imitated by competitors, whereas human resources and management capabilities signify internal core capabilities creating value (Bekmamedova & Shanks, 2014; Huang et al., 2018). In the augmented analytics concept, the role of human and managerial factors might increasingly be challenged in several respects. The most significant challenge is that augmented analytics will automate a large part of the activities and tasks throughout the analytics cycle with the help of AI, including data preparation, data analysis, and recognition of patterns and correlations to enable improved decision making (Prat, 2019). At first glance, this increasing level of automation in the analytics cycle is expected to decrease the relevance of human resources, implying that the role of data scientists might be less important once AI has taken over.

To gain insights from unstructured data, data scientists need to understand the entire data science cycle (Goyal et al., 2020). The most important tasks of data scientists include the creation of models that help to gain competitive advantage for their organization (Ahmed et al., 2019). However, extensive data growth in recent years has increased the complexity of analytics tasks. For example, due to difficulties in data access and time-consuming data preparation steps, data scientists typically focus on problem solving rather than tactical analysis processes and strategic tasks. As a result, they lack the time to develop and process complex analysis models to keep pace with economic changes (Nguyen, 2016, p. 22). According to a recent study, data scientists spend 60% of their time on non-value-adding, manual activities such as data collecting, cleaning, and organizing, rather than on core tasks such as creating models and algorithms (Earley, 2017). To reduce manual in favor of analysis tasks, augmented analytics uses ML to automate data preparation, discovery, and knowledge sharing for a wide range of business users, operational staff, and data scientists (Prat, 2019). Abas et al. (2020) forecast that 50% of the analytics workflow will be automated with augmented analytics across industries by 2022. From 2028 onwards, the focus will finally be on the large-scale application of augmented analytics. Although the entire analysis process will be automated at that point, data scientists will still be required to review and validate the results, and to perform higher order tasks (Abas et al., 2020). Data scientists are thus not made redundant; rather, they benefit from augmented analytics, as they can spend their time on more value-adding tasks (Prat, 2019). In addition, augmented analytics also constitutes a solution to the mentioned skills shortage in the data science market (LaPlante, 2019; Prat, 2019). While analytics tasks have become more complex with the extensive data growth in recent years, the skills gap has continued to widen (Cao, 2018, p. 130). Finding qualified and skilled analytics personnel is still a major challenge for firms (Grover et al., 2018; Mikalef et al., 2018); hence, augmented analytics provides firms with the possibility to develop self-learning and self-optimizing models with less involvement from human analysts (Davenport and Harris, 2017; LaPlante, 2019; Prat, 2019).
In addition to the major impact on the human factor of the socio-technical system, augmented analytics is expected to change the role of management capabilities. In the BDA literature, the term “management capabilities” describes the ability to develop competitive actions and make solid business decisions based on insights gained from data (Akter et al., 2016; Anand et al., 2016). Since data-based acquired knowledge by itself is of no value (Božič & Dimovski, 2019b), adequate managerial efforts in acquiring and analyzing critical information for decision making are essential to create value from BDA (Anand et al., 2016; Torres et al., 2018). Augmented analytics therefore allows so-called “citizen data scientists” such as business users or analysts to easily access advanced analytics (Prat, 2019). While in the past, only skilled data scientists were able to create models, the increasing automation of data science tasks will enable future citizen data scientists to produce far more advanced analyses (Cearley et al., 2017). Thus, the automated nature of augmented analytics enables business users and executives to make faster decisions and gain competitive advantage through acquiring relevant insights from data without being necessarily dependent on data scientists or other IT professionals (LaPlante, 2019).

5 Application Fields and Use Cases for Augmented Analytics

According to a recent report, augmented analytics addresses various end users from six industry sectors: banks, financial services, and insurance companies (BFSI); telecommunications and IT; healthcare; government; logistics; retail and others (Reportlinker, 2019a). Based on our qualitative and quantitative content analyses of the literature and Crunchbase data, we identified use cases associated to the following three industry sectors: BFSI, healthcare, and transportation and logistics. In the following, we describe these use cases in more detail.

5.1 Banks, financial services, and insurance

The BFSI segment is the most important user group of augmented analytics, accounting for approximately 25% of total sales in 2017. The major role of augmented analytics in this sector can be explained by the fact that more accurate forecasts bear an immense potential to reduce costs, increase revenue, and comply with regulations, which equally facilitates risk mitigation and improves fraud detection (PRNewswire, 2019). By using augmented analytics in BFSI, marketing strategies and customer loyalty policies can be improved, new investment strategies can be developed, and risks can be reduced (Reportlinker, 2019b). For example, augmented analytics supports BFSI in risk management by enabling continuous auditing and monitoring in several areas, including regulatory environment, supply chain, and business strategy (Huttunen et al., 2019). With continuous auditing and monitoring processes, augmented analytics can also contribute to the prevention of financial fraud (Huttunen et al., 2019).

Especially in the BFSI sector, financial fraud has become a serious problem with devastating economic consequences in recent years. Numerous cases are known, including bank fraud, insurance fraud, securities and commodities fraud (Ngai et al., 2011). The most well-known case is probably Wirecard’s accounting fraud that caused extensive financial damage through the fraudulent financial statements. Despite several warning signs of accounting malpractice as well as fraud allegations by the media in 2015, the fraudulent practices of Wirecards were detected no earlier than 2020 (O’Donnell, 2020; Zeranski and Sancak, 2020). Prior research already demonstrated the usefulness of data mining techniques based on algorithms such as decision trees, neural networks, and Bayesian networks for detecting fraudulent financial statements (Kirkos et al., 2007). In BFSI, ML-based fraud detection solutions can analyze past transaction data to identify fraudulent patterns in customer transactions and activities (Kashyap, 2017). In addition to the aforementioned regulatory reporting and fraud detection use cases, companies from the BFSI sector can develop and improve their products and services through augmented analytics. For example, based on continuous analysis of specific disease characteristics and patient data, the insurance industry can offer novel and appropriately customized health insurance policies to customers at low cost (Kale et al., 2020).
5.2 Healthcare

Among the main markets for augmented analytics, the healthcare sector is forecasted to experience strong growth potentials by 2025 due to the increasing demand for optimized care services and data-enabled insights for health executives and other stakeholders (PRNewswire, 2019). This growing demand is attributed to the rising population age and the increase in chronic diseases and insurance coverages. As a consequence, technologies are needed that help to manage healthcare processes more efficiently (Oesterreich et al., 2020). Especially AI-enabled technologies are expected to have a major impact on healthcare (Greaves et al., 2019). Recent forecasts reveal that the healthcare system can massively benefit from the use of big data and ML, with an annual revenue of $100 billion (Lamberti et al., 2019). Applicable in many healthcare settings, AI-enabled analytics can personalize and improve healthcare services through continuous patient monitoring and collection of their physical, personal, social, and medical data. This individualized concept is also known under the term “augmented personalized healthcare” (Sheth et al., 2017).

Furthermore, the use of big data and AI in healthcare makes sense wherever data-based insights are important. This is the case, to name a few, in disease diagnosis, drug development and manufacturing, and epidemic prediction (Lamberti et al., 2019). For example, a ML model that uses all available data on disease symptoms, incidence, potential treatments, behavioral aspects, and health trends could help physicians diagnose diseases faster and treat them more effectively (Dubey, 2020; Kashyap, 2017).

Currently, data on epidemic outbreaks can be collected from the Internet, with possibilities for real-time updates in social media and other sources, which an augmented system could use to continuously predict epidemic outbreaks (Dubey, 2020). For example, ML and NLP can be used in the current COVID-19 pandemic to more accurately and continuously predict future spread dynamics (Lalmuanawma et al., 2020; Raza, 2020) and develop appropriate countermeasures and strategies.

5.3 Transportation and logistics

Augmented analytics is considered worthwhile also in the transportation and logistics sector for several reasons. With the help of AI, a digital supply chain can be created in which data can continuously be analyzed to automate decision making and optimize the entire supply chain (Hellingrath and Lechtenberg, 2019). For transportation companies, accurate real-time predictions can help to optimize a number of activities and processes along the entire supply chain. These include selecting the best and most profitable transportation services, forecasting future incidents, providing timely feedback on transportation systems, and analyzing vehicle delays to optimize future predictions (Kashyap, 2017, p. 225).

Similarly, augmented analytics is associated with smart mobility, as it has the potential to automate road, rail, and air traffic. Public transport data such as national road traffic, road user behavior, or networked vehicles can provide IT and mobility experts with valuable insights on the basis of which road traffic can be developed or optimized (Trombin et al., 2020). In a smart infrastructure, automated data collection and analysis helps save fuel. Additionally, connected systems contribute to reducing traffic volumes and avoiding accidents and dangerous incidents (Kashyap, 2017, p. 135).

6 Discussion

6.1 Implications for research and practice

Following the call for more research on the recent emergence of BDA in the IS field in general (Baesens et al., 2016) and the high attention on augmented analytics as promising BDA concept in particular (Prat, 2019), our paper contributes to the body of knowledge by answering three main research questions. Additionally, we developed an augmented analytics framework based on the analytics cycle of Prat (2019) and the big data framework of Gupta et al. (2018) to consistently summarize the main findings and highlight possible avenues for future research. The proposed
framework combines the seven successive analytics steps (light blue boxes: I. Business understanding, II. Data preparation, III. Data analysis, IV. Model deployment, V. Decision, VI. Action and VII. Monitoring) adapted from Prat (2019) with the three-layer framework of Gupta et al. (2018) (dark blue arrows). We enriched this with our main findings on the role of human and organizational resources in each analytics step for gaining insights and creating value in the identified application fields and use cases. Thereby, we considered the enabling AI-based techniques such as ML primarily for analyzing structured and NLP for unstructured data. The framework in Figure 3 consists of two layers representing the resources and application fields in which augmented analytics can be of value for business and society.

The inner layer illustrates the role that technological, human, and organizational resources play at the different stages of the analytics cycle. In answering RQ1, we highlighted the potential of AI technologies such as ML and NLP to automate the analytics cycle (Prat, 2019). However, to date the limits in automation are met in terms of decision making and identification of business problems and opportunities (Prat, 2019). Although automated decision making using AI algorithms is technologically possible, managers would have to blindly rely on data quality and analytical processes. Yet, in most cases, humans have the final say when it comes to making decisions and taking action (Schmidt et al., 2020). Creating trust in automation and thus in applied AI relies on aspects related to the process, performance, and purpose levels of analytics (Hengstler et al., 2016). A black-box tool that generates results whose foundations are, however, neither comprehensible nor explained is hardly entrusted with strategic business decisions or, as in healthcare, patient safety decisions (Baesens et al., 2016). Further research is needed in the areas of transparency and explainable AI (Schmidt et al., 2020) as well as data governance (Prat, 2019) to ensure data quality and algorithms that allow trustworthy augmented analytics.

Figure 3. Framework of augmented analytics.

Guided by the IT business value model, we outlined the role of technological and human resources as well as management capabilities in the augmented analytics concept. To address RQ2, we delved deeper into the changes that the AI-driven nature of augmented analytics bring to its socio-technical
concept. The results reveal that the introduction of augmented analytics will produce permanent, significant changes to the human and organizational components of its socio-technical concept. As stated by Prat (2019), in the future even phases that to date are outside the scope of analytics tools, such as the initial phase of business understanding or the last phases of operational decision making and action taking, may become increasingly automated by employing ML models (e.g., in the case of high-frequency trading). As AI cannot take over the full spectrum of tasks and activities within the analytics cycle, human resources and management capabilities are still necessary for a variety of tasks in data analysis and model deployment as well as in the decision, action, and monitoring phase. Nevertheless, future research needs to address how the job scope of analytics personnel will change against this backdrop. Furthermore, the impact of the automation that AI brings to the analytics cycle on the job market is also an interesting question to be answered, including the emergence of the so-called “citizen data scientist” (Prat, 2019).

While the management role in analyzing critical information remains essential for decision making and creating value from data-based insights, the introduction of augmented analytics will cause a cultural shift away from opinion-based decisions to fact-based decisions. However, a vital precondition for an effective interplay between the technological, human, and managerial components of the augmented analytics concept is trust, which is especially true for AI-driven augmented analytics (Prat, 2019). Yet, despite being a key factor in data-enabled decision making, extant studies acknowledge that in BDA research trust has remained a rather under-examined area to date (Baesens et al., 2016). Decision makers will not benefit from data-driven insights when they do not trust the data or the underlying analytics techniques. Given the aforementioned black-box problem, trust is even more important in augmented than in non-AI-driven analytics, making transparency and understandability of analytics models essential requirements (Baesens et al., 2016). Trust, transparency, and credibility are therefore worthwhile future research avenues. Future research could examine the main antecedents of trust in augmented analytics and how augmented analytics must be designed to be trustworthy (Prat, 2019). Other meaningful research questions are whether, how and to what extent augmented analytics truly contributes to improved decision making (Prat, 2019). According to the theory of technology dominance, the reliance on intelligent decision support can, under certain conditions, have a long-term negative effect on the quality of the user’s decision (Arnold and Sutton, 1998). Hence, future research should investigate which conditions have profitable and which detrimental effects on decision making and quality. Similar to other BDA concepts (Kohli and Grover, 2008), the question of how, when, and why augmented analytics creates value for business users is also a fundamental issue for IS researchers.

Regarding **RQ3**, we identified three main application fields of augmented analytics represented by the outer layer of the framework in Figure 3. Among others, the application of augmented analytics is considered worthwhile in a broad range of use cases for the BFSI, healthcare, and transportation and logistics sectors. According to recent reports (Reportlinker, 2019a, 2019b), augmented analytics is even targeted at end users from other sectors such as telecommunications and IT, government, and retail. However, we only included the application fields and use cases in this framework identified through our analyses. To enable the application of augmented analytics in business practice, several ethical, legal, and social questions must be addressed. For example, the continuous analysis of data, as is required in the augmented analytics concept, raises questions concerning data security and privacy issues. In healthcare, for example, the continuous use of highly sensitive personal data for health and economic purposes may violate the privacy of patients and thus must conform to legal or ethical requirements when being used in business practice (Oesterreich et al., 2020). The investigation of these ethical and legal questions concerning data security and privacy issues of data use, regardless of the application field, constitutes another reasonable avenue for future research.

Overall, our research contributes to a better understanding of the augmented analytics concept by addressing open questions regarding its socio-technical system, including technologies, people, processes, and their interactions (Prat, 2019). Given the fact that the respective research field is still considered still in its early stage (Prat, 2019), our insights greatly expand the body of knowledge. Due to the interdisciplinary nature of the BDA research field, scholars from across diverse disciplines,
including IS research and related fields of computer science, marketing, management, communication, and mathematics, can benefit from our study. In addition, we uncover prevalent research gaps and highlight future research directions. For practitioners, the findings are valuable for gaining a practice-oriented understanding of augmented analytics, its conceptual characteristics, the technological, human, and organizational resources that are required to use augmented analytics in business practice as well as potential application areas and use cases. Based on this understanding, they can decide whether the adoption of augmented analytics is worthwhile for their business practice and whether the required resources are available in their organization.

6.2 Limitations

Our research paper has several limitations that must be considered when interpreting the results. The first limitation concerns its focus on augmented analytics, a relatively new topic of interest. Therefore, the lack of high quality refereed articles that can serve as valid basis for our qualitative and quantitative content is one of the main limitations. As we decided to select as many relevant articles as possible to broaden the literature basis, we also included practical publications such as newspapers, magazine articles, white papers, and reports as well as working papers in the final sample for the exploratory analysis. Thus, our insights are not exclusively academic; practice-oriented insights into augmented analytics are also provided, including a market analysis of augmented analytics solutions. Accordingly, a potential bias may exist from the inclusion of non-reviewed articles and their possibly varying qualities.

The second limitation refers to the potential selection bias arising from the assessment step within the systematic literature search. Although conducted independently by two researchers with an IRA of 0.79, we cannot guarantee that our literature selection is completely free from selection bias. The same applies for the coding step for the qualitative content analysis to inform the concept matrix. The insights gained in the qualitative content analysis step might also be biased in the sense that important and relevant insights might have been overlooked. Another limitation lies in the applied text mining method to quantitatively examine the content of the selected articles. Although text mining can help researchers to draw “replicable and valid inferences” based on the text corpus (Krippendorff, 1989, p. 403) by performing text analytical procedures to identify recurring text patterns in the literature sample (Roberts, 1997), the conclusions drawn from the findings remain restricted to the underlying secondary data. The selected text analytical procedures were conducted using the R programming language and dedicated text mining and visualization packages (e.g., quanteda, stm, spaceyr, RTextTools, and ggplot). As a consequence, the findings are based on the analytical outputs of these procedures and might have been different when applying other tools or text mining packages.

7 Conclusion

In recent years, augmented analytics has increasingly attracted the attention of researchers and practitioners as one of the more advanced, novel approaches for handling big data (Davenport & Harris, 2017; Prat, 2019). Due to the novelty of this concept, the surrounding research field is still in its early stages, with many open questions regarding its socio-technical system, that is, technologies, people, processes, and their interactions (Prat, 2019). Responding to this gap, we examine the role of the technological, human, and organizational factors within the analytics cycle as well as the main use cases in which augmented analytics can be of value for business and society. By employing an exploratory research approach based on quantitative and qualitative content analyses, we intend to enhance the understanding of augmented analytics and its implications for research and practice. The results indicate that the introduction of augmented analytics will have significant impacts on human and organizational issues, with the role of analytics personnel and management being increasingly important. Although augmented analytics is praised to offer several benefits compared to conventional BDA techniques, numerous questions concerning its use remain open. The findings of this paper contribute to the body of knowledge by enhancing the understanding of the augmented analytics concept, uncovering some prevalent research gaps, and highlighting future research directions.
References


