



Revista de Administração Contemporânea

ISSN: 1415-6555

ISSN: 1982-7849

Associação Nacional de Pós-Graduação e Pesquisa em Administração

Araujo da Fonseca, André Luis; Castro Pires de
Souza Chimenti, Paula; Suarez, Maribel Carvalho
Using Deep Learning Language Models as Scaffolding Tools in Interpretive Research
Revista de Administração Contemporânea, vol. 27, no. 3, e230021, 2023
Associação Nacional de Pós-Graduação e Pesquisa em Administração

DOI: <https://doi.org/10.1590/1982-7849rac2023230021.en>

Available in: <https://www.redalyc.org/articulo.oa?id=84075233008>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UABM  redalyc.org

Scientific Information System Redalyc
Network of Scientific Journals from Latin America and the Caribbean, Spain and
Portugal

Project academic non-profit, developed under the open access initiative

Invited Article

Using Deep Learning Language Models as Scaffolding Tools in Interpretive Research

Usando Modelos de Linguagem Baseados em Aprendizado Profundo como Ferramentas de Auxílio em Pesquisa Interpretativa



André Luis Araujo da Fonseca¹ 
Paula Castro Pires de Souza Chiment¹ 
Maribel Carvalho Suarez¹ 

■ ABSTRACT

Objective: the paper introduces a framework for conducting interpretive research using deep learning algorithms that blur the boundaries between qualitative and quantitative approaches. The work evidences how research might benefit from an integrated approach that uses computational tools to overcome traditional limitations. **Proposal:** the increased availability and diversity of data raises the utility of algorithms as research tools for social scientists. Furthermore, tuning and using such computational artifacts may benefit from interpretive procedures. Such circumstances turn the traditional debate between quantitative and qualitative research on its head: the research strategy that likely yields the most assertiveness and rigor is the one that may require vigorous hermeneutic effort. Along these lines, neural word embeddings can be instrumental in allowing researchers to read the data closely before and after interpretation.

Conclusions: to take advantage of the opportunity generated by these new algorithms, researchers may broaden their previous conceptions and adopt a participative point of view. In the coming decades, the interweaving of computational and interpretive methods has the potential to integrate rigorous social science research.

Keywords: qualitative research; language models; deep learning; epistemology.

■ RESUMO

Objetivo: o artigo apresenta um *framework* para a condução de pesquisa interpretativa usando algoritmos de aprendizado profundo que dissolvem os limites entre abordagens qualitativas e quantitativas. O trabalho mostra como a pesquisa pode se beneficiar de uma abordagem integrada que usa ferramentas computacionais para superar limitações tradicionais. **Proposta:** a maior disponibilidade e diversidade de dados tem aumentado o valor dos algoritmos como ferramentas de pesquisa que podem ser utilizadas pelos cientistas sociais. Além disso, ajustar e usar esses artefatos computacionais pode requerer e se beneficiar de procedimentos interpretativos. Tais circunstâncias viram de cabeça para baixo o tradicional debate sobre pesquisa quantitativa e qualitativa: a estratégia de pesquisa que provavelmente produz a maior quantidade de assertividade e rigor é aquela que requer esforço interpretativo mais vigoroso. Nesses termos, a codificação de palavras por meio de redes neurais pode ajudar os pesquisadores a ler com mais proximidade os dados, antes e depois de interpretá-los. **Conclusões:** para aproveitar a oportunidade produzida por esses novos algoritmos, pesquisadores poderiam ampliar as concepções existentes e adotar um ponto de vista epistemológico mais participativo e agnóstico. O entrelaçamento de métodos computacionais e interpretativos tem potencial de integrar e viabilizar interpretação e mensuração com rigor analítico nos contextos de abundância de dados que devem caracterizar a pesquisa em ciências sociais nas próximas décadas.

Palavras-chave: pesquisa qualitativa; modelos de linguagem; aprendizado profundo; epistemologia.

* Corresponding Author.

1. Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração, Rio de Janeiro, RJ, Brazil.

Cite as: Fonseca, A. L. A., Chiment¹, P. C. P. S., & Suarez, M. C. (2023). Using deep learning language models as scaffolding tools in interpretive research. *Revista de Administração Contemporânea*, 27(3), e230021. <https://doi.org/10.1590/1982-7849rac2023230021.en>

JEL Code: M0.

Editor-in-chief: Marcelo de Souza Bispo (Universidade Federal da Paraíba, PPGA, Brazil) 

Published as Early Access: March 08, 2023.
Assigned to this issue: April 25, 2023.

"In interpretive research, BERT could potentially be used to analyze and interpret large amounts of text data, such as qualitative interview transcripts or social media posts. It could be used to identify patterns and trends in the data that might not be immediately apparent to a human researcher, or to generate summaries or extract key themes from the text data ..."

Answer given by the chatGPT language model to the question: "Do you believe it is a good idea to use BERT to facilitate interpretive research?" (December 23, 2022)

INTRODUCTION

Park et al. (2023) have recently published in *Nature* a paramount article showing the decline of innovation and disruption in science. The authors analyzed more than 45 million papers, noticing a substantive change in science and technology that reinforces concerns about slowing innovative activity. They attribute this trend partly to scientists' reliance on a narrower set of existing knowledge, with papers increasingly less likely to break with the past in ways that push science in new directions. This situation aligns with a period of normal science, accordingly to the definition proposed by Kuhn (Matthews, 2022), that we may be experiencing since researchers' beliefs and practices have become more pragmatic (Morgan, 2007). Given the greater complexity of the present-day phenomena and the structuring of academic careers, among other factors, the article mentioned above seems to herald a period of crisis in science, evidenced by stagnation. In this sense, it also highlights the urgency of new proposals for reformulating and adapting principles and methods.

Focusing on business and management research, we believe that scholars may face even more challenging situations in the future. Indeed, society's transformations demand changes in how managers run firms. To mention just one of the many aspects, such a requirement arises partly from stakeholders increasingly adopting a position in which businesses must contribute to contemporary issues (Friede et al., 2015; Scheyvens et al., 2016). However, the decision-making (epistemology) that can lead to this contribution depends on a growing and increasingly unstructured volume of data (El-Kassar & Singh, 2019). Therefore, management processes and toolboxes may need to change for organizations to bring such a desire to fruition (Sinkovics et al., 2021).

The observed lack of innovation in academic production combined with ongoing changes in its objects of study (businesses) represents one of the most critical issues researchers must address. This situation resembles the internet's early days when a portion of the interaction among consumers migrated to virtual spaces for the first time. As a result, a firm's efforts to gain marketing insights suddenly required analyzing data produced by individuals

on digital platforms (Godes & Mayzlin, 2004). The same circumstances affected management researchers. Their ability to explain consumer behavior also started to require collecting and analyzing vast amounts of digital data. Traditional interpretive research methods, such as ethnography, were expanded and adapted to cope with these novel aspects (Kozinets, 2002).

This transformation has accelerated in the last few years. The amount and variety of marketing-related digital data have significantly increased (Balducci & Marinova, 2018). Thus, it does not sound credible to claim that the explosion of new data over the last two decades has not limited management researchers' ability to observe and interpret relevant marketing information using solely old approaches (Berger et al., 2022; Erevelles et al., 2016). In addition, research methods previously adjusted to deal with virtual contexts must be updated again. This time, it may not be possible to merely incorporate the observation of new data types while maintaining a position in which human reasoning is the only thing to rely on (Kozinets et al., 2018). Indeed, it may be the case that the fundamental notion of observing a business-related process has to change, and such a reconceptualization may require the adoption of a more participative epistemology (Campagnolo, 2021).

Aligned with the necessity of reflecting on new methodologies, we suggest a potential disruption with traditional methods, blurring the boundaries between qualitative and quantitative approaches. We advocate that research (and researchers) might benefit from an integrated approach that uses computational tools to overcome traditional limitations. In this sense, we aim to contribute to the contemporary debate on research methods by anticipating disruptions, detaching from the present, and trying to understand the contemporary as the result of a historical process (Bispo, 2022).

We structure this paper as follows: in the next section, we outline the current status of the quantitative-qualitative debate. Next, we discuss deep learning language models as reading aids in qualitative research. Later, we propose a framework to integrate algorithmic artifacts in interpretative research and briefly discuss how those elements interact in such a situation. Then we present an actual research case using a language model to scaffold an interpretative research

effort. We close the article by discussing its contributions and opportunities for future research.

Fuzzies strike back: The quantitative-qualitative debate in an era of data abundance

Reviving the debate over quantitative versus qualitative research near the end of the first quarter of the twenty-first century risks instilling readers with a sense of boredom. Indeed, social scientists and philosophers have discussed this subject with less or more energy for several decades (Bryman, 1984). As a result, the relationship between these two ways of researching social phenomena has changed over time, going through periods of conflict and cooperation (Smith & Heshusius, 1986). A recent empirical article demonstrates that commonplace beliefs in the existence of entirely distinct methodological cultures, quantitative and qualitative, between which communication is supposed to be difficult (Goertz & Mahoney, 2012), are somewhat incorrect (Kuehn & Rohlfing, 2022).

Rapid social change rejuvenates the game field in which the quantitative-qualitative debate occurs. Social science researchers must all contend with an explosion of new data in terms of volume and heterogeneity as new digital sources provide “rich detail about the evolution of social relationships across large populations as they unfold” (Edelmann et al., 2020, p. 2). Indeed, the unmeasurable was turned into measurable by “the technological revolution in mobile, web, and internet communications” (Watts, 2011, p. 266). Such change “has the potential to revolutionize our understanding of ourselves and how we interact” (Watts, 2011, p. 266), amongst other elements.

Thus, we argue that the societal transformation described above fuels discussions and redefinitions about rigor in social science research (Sells et al., 1995) for all practical matters. Closed-ended surveys, an example of a popular tool in quantitative-deductive inquiry, allow for progressively limited investigations if used alone. Indeed, scholarship work may become more inductive (Grimmer et al., 2021) and unobtrusive in comparison to asking individuals to answer a set of predefined questions in a form due to the data’s permanent state of *flow* and researchers’ interest in investigating each phenomenon with the maximum possible degree of realism (Chang et al., 2014). In addition, data abundance and complexity make it less practical or desirable to organize information in a conventional quantitative manner, with variables in columns and observations in rows (Lazer et al., 2020). In sum, the increasing intricacy of social phenomena and the abundance of “unfamiliar variables and data formats” (Brandt & Timmermans, 2021, p. 191) may

make it less practical the utilization of hypotheses-testing approaches.

That being the case, machine learning algorithms (Borch, 2021) may be needed to identify hidden concepts in phenomena-related massive heterogeneous datasets (Lazer et al., 2021). As a result, it seems plausible to conjecture that a portion of the rigor of future social science research may derive from the appropriate selection and usage of AI-related computational artifacts to scaffold both quantitative and qualitative work.

What matters most for the argument we aim to advance is that the transformations listed above may represent an opportunity for interpretive researchers, individuals commonly associated with the term *fuzzy*, as Hartley (2017) proposed. Similar to what happened with the designers of self-driving vehicles that had to ask for the help of anthropologists to “cope with the social and contextual complexity of mixed-traffic interaction” (Rothmüller et al., 2018, p. 482), the use of new algorithms in scholarship research on social sciences may create a larger space for qualitative reasoning. Indeed, when computational methods that can “provide much better predictive performance” than statistical inference methods (Grimmer et al., 2021, p. 398) are used to analyze large heterogeneous datasets, social science scholars face a dilemma. The mapping from data to the predictions made by such algorithms is not directly given as in a regression, where the estimated coefficients indicate the effects caused by the independent variables (Gentzkow et al., 2019). Thus, researchers’ ability to explain social phenomena by leveraging deep learning models, for example, may strongly depend on a theory-based interpretation of results produced by brute-force computation (Grimmer et al., 2022).

These transformations seem to be aligned with Campagnolo’s (2021) vision of participative epistemology (Heron & Reason, 1997) as “an empiricist notion of knowledge production whereby all those involved in the research endeavor (i.e., theories, methods, epistemic objects, and subjects) are seen as both co-researchers, whose agency contributes to generating ideas, designing and managing the project, and drawing conclusions; and co-subjects, participating in the activity that is being researched” (Campagnolo, 2021, p. 4). We claim that the increasing amount and complexity of data used in social sciences and the power of machine learning algorithms to uncover hidden relationships may end up generating a new way of producing knowledge in social sciences. In this sense, we follow Campagnolo’s call for transcending the current division between paradigms and research approaches (Campagnolo, 2021).

Substantive research increasingly benefits from an epistemology in which the engagement of researchers before

and after utilizing machine learning algorithms becomes necessary to produce helpful descriptions and explanations. According to [Grimmer et al. \(2021\)](#), the “current abundance of data allows us to break free from the deductive mindset that was previously necessitated by data scarcity” (p. 396). As a result, it seems likely that scholars and institutions

inevitably treat the quantitative-qualitative debate as an “internal and obsolete distinction” ([Campagnolo, 2021](#)).

Table 1 shows the differences between the four approaches discussed here: qualitative, quantitative, mixed, and participative.

Table 1. Approaches comparison.

	Qualitative	Quantitative	Mixed	Participative
Worldview	Constructivism; Interpretativism	Post-positivism	Pragmatism	Multiple
Data collection methods	Cases, subject selection, small sets of data deeply observed; Open-ended questions; Emerging approaches; Text or image data	Surveys, large sets of data, correlations, patterns; Closed-ended questions; Predetermined approaches; Numeric data;	Qualitative and quantitative data	Structured and unstructured qualitative and quantitative data
Data analysis	Hermeneutic, coding, human analysis	Software-aided analysis	Sequential; Concurrent; Transformative	Rhizomatic
Research designs	Grounded theory; Ethnography; Case study; Narrative; Phenomenology	Surveys; Experiments		Computational grounded theory; Extended computational case study; Cultural cartography
Ontology	Subjective: reality is socially built	Objective: reality is the empirical world	Subjective and objective	Subjective, Objective, and virtual
Epistemology	Subjective: researcher is part of the process and influences its outcomes	Objective: the correspondence between our statements and reality through inductive verification or deductive falsification	Subjective and dependent on practical consequences	Temporary joint epistemic work; Abductive reasoning

Note. Source: Elaborated by the authors based on [Creswell \(2013\)](#), [Van de Ven \(2007\)](#), [Deleuze and Guattari \(1980\)](#), and [Campagnolo \(2021\)](#).

The table summarizes some elements that we develop throughout this article. The first is that the proposed data analysis is neither sequential nor concurrent; it is rhizomatic. That is, it is not hierarchical and has no fixed order. Each step along the research path can lead to any other portion of the same course.

The second concerns ontology, which is at the same time subjective (reality is constructed), objective (reality is what the data show), and virtual (reality is what emerges from the execution of algorithms). The idea of virtual comes from the work of [Deleuze \(1968\)](#): ‘the virtual is real’.

The third point refers to abductive epistemology. After analyzing the algorithms, the researcher asks himself: What could have made these algorithms tell me this? In such

cases, without algorithms, the researcher would not see what they produce and, therefore, would not make such logical inferences.

In addition, the approach is participative, meaning that all forms of knowledge construction are involved in generating conclusions, followed by abductive reflection.

Thus, provided that the widespread presence of computational elements in society expands the possibilities of academic research and modifies the social processes one wishes to study, we follow [Marres \(2020\)](#) in saying that algorithmic approaches to research greatly benefit from interpretative procedures and vice versa. All data types, particularly textual ones, can now be ‘read’ in “its full hermeneutic complexity and nuance” ([Mohr et al., 2015, p. 3](#)).

Deep learning language models as reading aids in qualitative research

“Deep learning text models are a type of neural network model that are specifically designed to process and understand natural language text. These models can be used for a wide range of natural language processing tasks, such as language translation, text summarization, sentiment analysis, and question answering. These models are usually trained on large amounts of text data and are able to learn to understand and generate human language.”

Answer given by the chatGPT language model to the question: What are deep learning text models? (January 17, 2023)

Computational methods are ‘observation and reading tools’ (Mohr et al., 2020). They may assist scholars in perceiving and incorporating into their hermeneutics patterns that are difficult to visualize, such as those connected to the less accessible informants in an ethnographic study (Campagnolo, 2021). For example, algorithms can enhance interpretative research, analyzing with accuracy large amounts of text data, such as interviews or social media posts. Algorithms can be helpful during coding processes, identifying patterns or themes in the data. Thus, questioning whether a researcher would profit from learning to use word embedding models like Word2Vec (Mikolov et al., 2013) and BERT (Devlin et al., 2019), for example, corresponds to asking how well he wishes to ‘read’ in contrast to others. Although it may seem so, there is no exaggeration in the above statement. The truth is that the world’s increasing complexity recommends adopting such methods in qualitative research. While the area of ethnography, for example, is rich in epistemic diversity (Abramson et al., 2018), even an academic who strictly views his work as a “humanist enterprise within the general field of cultural production” (Abramson et al., 2018, p. 257) may enormously benefit from the new possibilities offered by those algorithms. Indeed, Marcus (1995) warns us that “any ethnography of a cultural formation in the world system is also an ethnography of the system” (Marcus, 1995, p. 99), which means that it examines “the circulation of cultural meanings, objects, and identities in diffuse time-space” (Marcus, 1995, p. 96). Thus, an ethnographer’s reading and observation abilities must be capable of dealing with the ‘rhizomatic’ (Deleuze & Guattari, 1980), in the sense of not being hierarchical, fixed, and stable, conditions inherent in modern social systems. That includes various contexts and circumstances involved so dynamically with each cultural setting that it is sometimes challenging for the researcher to observe and read everything. Why would a scholar attempt to deal with such diversity without utilizing those newly

available computational tools? What epistemic cost, if any, do many researchers associate with using algorithms?

Integrating algorithms in interpretative research: A framework proposal

For decades, the goal of automated textual content analysis has essentially been to read “large textual corpora in such a way that critical bits of information could be extracted and a measure of informational reliability could be calibrated” (Mohr et al., 2015). This approach is about extracting only the essential ideas in a text, precisely those with little room for argument regarding their existence (Mohr et al., 2015). Such characteristic means that these methods commonly disregard the “subtleties of expression, the complexities of phrasing and the more nuanced meanings of text corpora” (Mohr et al., 2015, p. 2). Because of this limitation, it is not surprising that researchers interested in executing a ‘close reading’ (Gavin et al., 2019) of the texts involved in each phenomenon have not historically shown much enthusiasm for this way of doing social science research.

However, access to new forms of data, such as transactional purchase data, records of social media usage, and data from public and private surveillance, has considerably increased the volume and variety of ‘texts’ that need to be read to produce insights into social science (Meckin, 2021). Due to this, algorithms such as BERT, which makes it feasible to apply previously acquired learning to interpret new text sentences, can make a massive difference in interpretive research. These algorithms can execute the coding of interviews, for example, “at a fraction of the effort of human-only coding while improving reliability” (Li et al., 2021, p. 1). Indeed, the explosion of data makes it expensive and difficult to ‘close read’ the totality of crucial ‘texts’ in each research situation. Luckily, deep learning algorithms may help reduce, amongst other elements, the number of dimensions in textual data, thus facilitating the interpretative process (Grimmer et al., 2021).

Because of the possibilities outlined above, we propose a framework for interpretive research in which algorithms and hermeneutic procedures are integrated. By doing this, we generalize the three-stage computational ground theory framework proposed by Nelson (2020). According to the notion that we want to advance, computational tools help researchers account for the multifaceted characteristics of modern social science phenomena. Thus, for researchers to have a better chance of answering “contemporary data-rich questions” (Nelson, 2020, p. 34), their method of conducting interpretive social science research should have the characteristics of an assemblage (Deleuze & Guattari, 1980) capable of producing a significant number of research paths using both algorithms and interpretative tasks. Figure

1 depicts a ‘research-as-assemblage-framework’ containing two spaces of possibilities, hermeneutic and algorithmic, whose elements interact in a multidirectional way and can be activated by researchers without a predetermined order. Such activations originate actual research paths.

Furthermore, as illustrated in Figure 1 in a non-exhaustive manner, this characterization of interpretive research as an assemblage includes several other elements. It

encompasses, for example, the coders who design and develop the algorithms and their beliefs and models. It also contains the data used to train the algorithms, which play an essential role in shaping their outputs. The precise balance between interpretation and computation in each research path is determined by the factors discussed by [Campagnolo \(2021\)](#). In the following section, we define an algorithm to help clear up any misconceptions preventing interpretive researchers from embracing such tools.

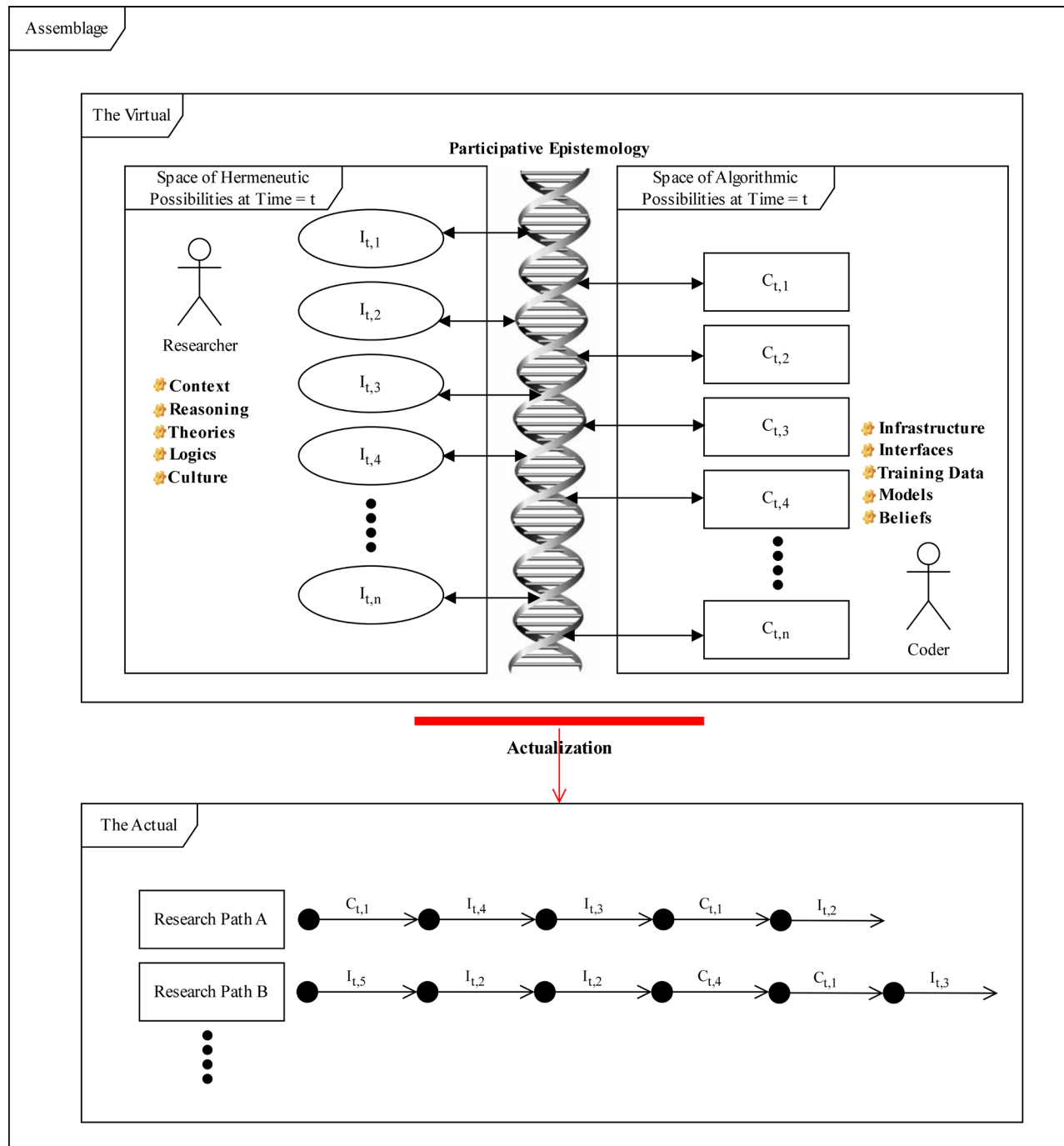


Figure 1. A ‘research-as-assemblage-framework’ for interpretive research in an era of data abundance.

Hermeneutic tasks and algorithmic procedures interact epistemically. What the researcher interprets/learns serves as input for the execution of algorithms and vice versa. Source: Elaborated by the authors.

Who is afraid of the big bad algorithm?

In recent years, the term ‘algorithm’ has been used to refer to various items that were formerly grouped under ‘software’ (Matzner, 2022). Scholars commonly take advantage of such a broad concept to talk about technological phenomena, such as the alleged influence of some computational systems on the “structure of public discourse or the functioning of state power” (Matzner, 2022, p. 2). This imprecise and abstract notion of an algorithm may lead interpretive scholars to ascribe to its use an epistemological cost that is barely grounded in reality.

Indeed, a scholar may associate using an algorithm in qualitative research with surrendering one’s epistemic beliefs to a “backstage device” that is responsible for “the constant optimization of experience” and undertakes forms of “infrastructural surveillance” (Cellard, 2022, p. 983). In these terms, it should be no surprise that ethnographers, for example, are often not very enthusiastic about using algorithms, especially those associated with deep learning. How would someone who views the research process as aiming for an “in-depth description of the phenomenon from the perspective of the people involved” (Yilmaz, 2013, p. 312) positively evaluate the use of a tool whose nature is currently associated with the construction of a social reality that disregards and possibly manipulates human will (O’Neil, 2016)?

The academic literature’s casual usage of the term ‘algorithm’ has resulted in a very non-technical understanding, making it a pejorative “catchy word symbolizing an opacity and growing dehumanization” (Cellard, 2022, p. 983). Indeed, we posit that the more fundamental definition of an algorithm as the sequence of steps required to complete a task (Christin, 2020) that is not always computational (Cellard, 2022) needs to be reinstated. Interpretive scholars should view an algorithm’s execution as an ‘assemblage’ (Siles et al., 2020) that encompasses the software being run in a computer (non-human) and a chain of human actors that includes each author of the code and the individuals who take advantage of the outputs, that is, themselves (Christin, 2020). Such an assemblage does not exist in a social vacuum. Because of that, Amore (2020) argues that answering the ethical question of whether algorithms are good or bad depends on how we use their signals. Instead of criticizing the opacity of deep learning algorithms, she advocates for a larger ‘aperture of observation’ in the decision-making processes that are supported by such assemblages. Such a notion means observing how the moments when the algorithm is wrong “give accounts of algorithmic reason” (Amore, 2020, p. 167).

Due to the impossibility of assigning an ethical value to algorithmic decisions, we follow Matzner (2022) in

suggesting that researchers commonly embrace a relational account of algorithms to consider each of them “an abstraction that is inherently related to a necessary complement of that abstraction in something concrete and particular” (Matzner, 2022, p. 2). In other words, “algorithms need to be seen as becoming what they ‘are’ only in relations” (Matzner, 2022, p. 3). The data that an interpretive scholar decides to feed the algorithm is one of these complementary relations. The second is the ‘aperture of observation’ through which the algorithm’s output influences the scholar’s interpretation. Thus, it is necessary to dispel the notion that deep learning algorithms possess an agency that would compete with the researcher’s agency. As we have said in this article, societal changes have pushed the amount of data available beyond the threshold that can be treated with the naked eye. While it is true that the employment of algorithms generates new ethical, ontological, and epistemological concerns (Christin, 2020), it is also true that they confer new capabilities on interpretive scholars. We believe developing a participative mindset is the key to taking advantage of these opportunities. Keeping an open and dogma-free mind, a crucial aspect of exploratory and inductive research, becomes increasingly important for developing substantive social science research that makes a difference.

Studying technology adoption with the help of a deep learning language model

In the following lines, we shall illustrate what it means to use machine learning in interpretive research by describing one of our most recent research projects. In work still in progress, we broaden the understanding of the continuous cycles of consumers’ enchantment and disenchantment in technology adoption (Belk et al., 2021). The first author’s research seeks to answer the following question: What produces, from a cultural point of view, the perpetuity of the desire to acquire each new version of a product whose current model may already meet consumer needs?

The project examines how current iPhone users express their feelings about new models on social media. Its goal is to compare the semantic distances between what they say and how Steve Jobs, the original product’s market introducer, described the first version. The differences in these distances are then interpreted using cultural theories on consumer desire. Indeed, it would be challenging to comprehend the phenomenon in all of its nuances without the possibility provided by computation. The semantic proximity of the two types of discourses provides strong empirical evidence that supports and expands on the interpretation produced by articulating the chosen theoretical lens. In addition, the first author’s understanding of the entrepreneurship phenomenon based on his experience and a dense reading of the same corpus, which the algorithm can scan to identify

patterns and themes, contributes to developing a new theorization.

The actual research is described briefly and shown in Figure 2. Its path corresponds to one of the infinite possible actualizations inherent in the assemblage of interpretive research that employs deep learning language models. The order of activation of each hermeneutic and computational possibility is determined by the researcher's desire and the project's circumstances. The interaction between these two spaces of possibilities can be described as a dialog with the following format: 'What I think -> What the algorithm says

-> What I think -> What the algorithm says', and so on. The epistemic process is participative in that it directly depends on all the elements being able to participate.

The dialogue concludes when the element who thinks last says enough for abductive reasoning to take place to produce a statement of the form: "Based on what we learned... it must be that..." In short, there are no set beginning or endpoints. The study concludes when the multiple interactions between the participating elements yield a sufficiently good explanation according to the criteria established inside the assemblage.

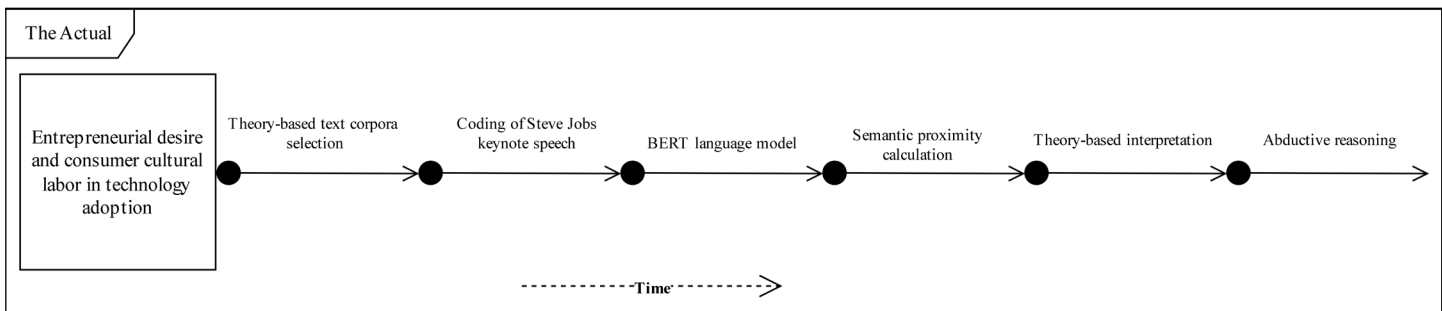


Figure 2. An interpretive study's research path using a deep learning language model as a scaffolding tool.

The research path goes from left to right to represent 'the passage of time'. While there is just one instance of every activity described here, our framework has absolutely no impediment (see Figure 1) for the repetition of algorithmic and interpretation tasks. In the general case, any activity can occur an arbitrary number of times at any point.

CONCLUSIONS

The proposed debate would be meaningless if not coherently aligned with relevant contemporary themes (Bispo, 2022). We believe that using advanced present-day computational tools in qualitative research represents an opportunity for social science scholars from developing countries. Aside from allowing researchers to understand and explain phenomena in a way that sometimes is impossible otherwise, algorithmic methods contribute positively to various operational aspects of theory-generating research (Nelson, 2020). They can, for example, help reduce the costs associated with coding textual content (Rodriguez & Storer, 2020). They also make it more practical and scalable to handle vast amounts of heterogeneous data obtained from various sources (Evans, 2014). In addition, computational methods can allow qualitative research in the *Global South* (Dados & Connell, 2012) to benefit from datasets originating in developed areas more effectively and accomplish the breadth of analysis that was previously only achievable in contexts or circumstances with higher budgetary abundance. This last aspect means that incorporating computational tools can help lessen firepower inequality in the global landscape of qualitative social science research.

Using computational methods in qualitative research may increase overall productivity (Nelson et al., 2018). This benefit should suffice in justifying their adoption, given the difficulty of undertaking substantive social research in countries like Brazil. However, we posit that such an element is not the primary reason why an interpretivist, for example, needs to consider employing algorithms in his work. Indeed, deep learning algorithms play an ever-increasing role in modern society. Text models such as chatGPT, which was recently introduced, have the potential to completely transform content-generating professions such as teaching and journalism (Haque et al., 2022). Instead of arguing that algorithms such as chatGPT reduce the value of a researcher's thinking to the point that they may pose a risk, this article seeks to advance a different type of insight.

The chatGPT, derived from the GPT-3 algorithm developed by the company Open.ai, can be used by social science researchers as a type of co-author with whom they can converse. This article talks about applying an equally powerful algorithm: BERT. It enables the first author to identify semantically similar elements of a 2011 keynote speech in the discourse of present-day consumers. From this identification, he derives his interpretation. Although the dialogue with BERT was not conducted in natural language, the outcome was comparable: BERT enabled us to read vast

amounts of cultural texts produced by a particular group of individuals in greater detail.

Business and management research may benefit from adopting a more participative epistemology. Marketers, for example, already use conventional data mining tools to generate insights. However, the present-day scholarship has not contributed significantly to that. Adopting deep learning language models as components of the research process corresponds to a window of opportunity that can enhance researchers' capacity to address contemporary issues by properly informing business practices. Conversely, if traditional epistemological silos are maintained, and researchers decide not to expand their toolboxes, there is a risk that marketing research will not adequately explain the world of contemporary consumers, widening the relevance gap (Starkey & Madan, 2001).

The importance of a researcher's ability to interpret is multiplied many times by algorithms such as chatGPT, BERT, and others. In other words, we argue that these models help capture the more general and quantifiable relationships. What remains to be done is primarily contextual and incommensurable—for instance, that which depends on local cultural elements. If the researcher adopts a participatory epistemological stance incorporating what the algorithm has learned as his or her learning, the dialogue with such computational artifacts shortens the distance separating the research and what is more substantive while assuring a high level of accuracy in the predictions made. In sum, deep learning language models are not only fuzzies' best friends but may also increase the ability of social scientists as a whole to conduct research that can benefit society.

REFERENCES

- Abramson, C. M., Joslyn, J., Rendle, K. A., Garrett, S. B., & Dohan, D. (2018). The promises of computational ethnography: Improving transparency, replicability, and validity for realist approaches to ethnographic analysis. *Ethnography*, 19(2), 254–284. <https://doi.org/10.1177/1466138117725340>
- Amoore, L. (2020). *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press.
- Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590. <https://doi.org/10.1007/s11747-018-0581-x>
- Belk, R., Weijo, H., & Kozinets, R. V. (2021). Enchantment and perpetual desire: Theorizing disenchanted enchantment and technology adoption. *Marketing Theory*, 21(1), 25–52. <https://doi.org/10.1177/1470593120961461>
- Berger, J., Packard, G., Boghrati, R., Hsu, M., Humphreys, A., Luangrath, A., Moore, S., Nave, G., Olivola, C., & Rocklage, M. (2022). Wisdom from words: marketing insights from text. *Marketing Letters*. <https://doi.org/10.1007/s11002-022-09635-6>
- Bispo, M. de S. (2022). Reflecting on contemporary Administration. *Revista de Administração Contemporânea*, 26(1). <https://doi.org/10.1590/1982-7849rac2022210203.en>
- Borch, C. (2021). Machine learning and social theory: Collective machine behaviour in algorithmic trading. *European Journal of Social Theory*, 25(4). <https://doi.org/10.1177/13684310211056010>
- Brandt, P., & Timmermans, S. (2021). Abductive Logic of Inquiry for Quantitative Research in the Digital Age. *Sociological Science*, 8, 191–210. <https://doi.org/10.15195/V8.A10>
- Bryman, A. (1984). The debate about quantitative and qualitative research: a question of method or epistemology? *The British Journal of Sociology*, 35(1), 75–92. <https://www.jstor.org/stable/590553>
- Campagnolo, G. M. (2021). Participative epistemology in social data science: Combining ethnography with computational and statistical approaches. *International Journal of Social Research Methodology*, 25(3), 1–13. <https://doi.org/10.1080/13645579.2021.1892379>
- Cellard, L. (2022). Algorithms as figures: Towards a post-digital ethnography of algorithmic contexts. *New Media & Society*, 24(4), 982–1000. <https://doi.org/10.1177/14614448221079032>
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80. <https://doi.org/10.1016/j.dss.2013.08.008>
- Christin, A. (2020). The ethnographer and the algorithm: beyond the black box. *Theory and Society*, 49(5–6), 897–918. <https://doi.org/10.1007/s11186-020-09411-3>
- Creswell, J. W. (2013). The Selection of a Research Approach. In *Research design: Qualitative, quantitative and mixed methods approaches* (pp. 30–55). SAGE Publications.
- Dados, N., & Connell, R. (2012). *The Global South. Contexts*, 11(1), 12–13. <https://doi.org/10.1177/1536504212436479>
- Deleuze, G. (1968). *Difference and Repetition*. Columbia University Press.
- Deleuze, G., & Guattari, F. (1980). *A thousand plateaus: Capitalism and schizophrenia*. University of Minnesota Press. <https://doi.org/10.4324/9780203584200>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. <http://arxiv.org/abs/1810.04805>

- Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, 46, 61-81. <https://doi.org/10.1146/annurev-soc-121919-054621>
- El-Kassar, A. N., & Singh, S. K. (2019). Green innovation and organizational performance: The influence of big data and the moderating role of management commitment and HR practices. *Technological Forecasting and Social Change*, 144, 483–498. <https://doi.org/10.1016/j.techfore.2017.12.016>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Evans, M. S. (2014). A computational approach to qualitative analysis in large textual datasets. *PLoS ONE*, 9(2). <https://doi.org/10.1371/journal.pone.0087908>
- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, 5(4), 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Gavin, M., Jennings, C., Kersey, L., & Pasanek, B. (2019). Spaces of Meaning: Conceptual History, Vector Semantics and Close Reading. In *Debates in the Digital Humanities 2019* (pp. 243–267). University of Minnesota Press. <https://doi.org/10.5749/j.ctvq251hk.24>
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574. <https://doi.org/10.1257/jel.20181020>
- Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4). <https://doi.org/10.1287/mksc.1040.0071>
- Goertz, G., & Mahoney, J. (2012). A tale of two cultures: Qualitative and quantitative research in the social sciences. Princeton University Press.
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for Social Science: An agnostic approach. *Annual Review of Political Science*, 24, 395–419. <https://doi.org/10.1146/annurev-polisci-053119>
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as data: A new framework for machine learning and the Social Sciences. Princeton University Press.
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). “I think this is the most disruptive technology”: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. <http://arxiv.org/abs/2212.05856>
- Hartley, S. (2017). Fuzzy and the techie - Why the liberal arts will rule the digital world. *Journal of Chemical Information and Modeling*, (9). <https://doi.org/10.1017/CBO9781107415324.004>
- Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, 3(3), 274–294. <https://doi.org/10.1177/107780049700300302>
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39, 61–72. <https://preventviolenceextremism.info/sites/default/files/The%20Field%20Behind%20the%20Screen-%20Using%20Netnography%20For%20Marketing%20Research%20in%20Online%20Communities.pdf>
- Kozinets, R. V., Scaraboto, D., & Parmentier, M. A. (2018). Evolving netnography: How brand auto-netnography, a netnographic sensibility, and more-than-human netnography can transform your research. *Journal of Marketing Management*, 34(3-4), 231–242. <https://doi.org/10.1080/0267257X.2018.1446488>
- Kuehn, D., & Rohlfing, I. (2022). Do quantitative and qualitative research reflect two distinct cultures? An empirical analysis of 180 articles suggests “no.” *Sociological Methods & Research*. <https://doi.org/10.1177/00491241221082597>
- Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, 595(7866), 189–196. <https://doi.org/10.1038/s41586-021-03660-7>
- Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., Wagner, C., & Lane, J. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062. <https://doi.org/10.1177/0002716217740116>
- Li, Z., Dohan, D., & Abramson, C. M. (2021). Qualitative coding in the computational era: A hybrid approach to improve reliability and reduce effort for coding ethnographic interviews. *Socius*, 7, 1–4. <https://doi.org/10.1177/23780231211062345>
- Marcus, G. E. (1995). Ethnography in/of the World System: The emergence of multi-sited ethnography. *Annual Review of Anthropology*, 24(1995), 95–117. <https://doi.org/10.1146/annurev.an.24.100195.000523>
- Marres, N. (2020). For a situational analytics: An interpretative methodology for the study of situations in computational settings. *Big Data and Society*, 7(2). <https://doi.org/10.1177/2053951720949571>
- Matthews, M. R. (2022). Thomas Kuhn and Science Education: Learning from the past and the importance of History and Philosophy of science. *Science and Education*. <https://doi.org/10.1007/s11191-022-00408-1>
- Matzner, T. (2022). Algorithms as complementary abstractions. *New Media and Society*. <https://doi.org/10.1177/14614448221078604>
- Meckin, R. (2021). Interdisciplinary report Computational Social Science: A thematic review. <https://eprints.ncrm.ac.uk/id/eprint/4476>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. <http://arxiv.org/abs/1301.3781>

- Mohr, J. W., Bail, C. A. B., Frye, M., Lena, J. C., Lizardo, O., McDonnell, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020). *Measuring culture*. Columbia University Press.
- Mohr, J. W., Wagner-Pacifci, R., & Breiger, R. L. (2015). Toward a computational hermeneutics. *Big Data and Society*, 2(2). <https://doi.org/10.1177/2053951715613809>
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48–76. <https://doi.org/10.1177/2345678906292462>
- Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods and Research*, 49(1), 3–42. <https://doi.org/10.1177/0049124117729703>
- Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods and Research*, 50(1). <https://doi.org/10.1177/0049124118769114>
- O'Neil, C. (2016). *Weapons of Math destruction: How Big Data increases inequality and threatens democracy*. Crown.
- Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942), 138–144. <https://doi.org/10.1038/s41586-022-05543-x>
- Rodriguez, M. Y., & Storer, H. (2020). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services*, 38(1), 54–86. <https://doi.org/10.1080/15228835.2019.1616350>
- Rothmüller, M., Rasmussen, P. H., & Vendelbo-larsen, S. A. (2018). Designing for interactions with automated vehicles: Ethnography at the boundary of quantitative-data-driven disciplines. *EPIC – Ethnographic Praxis in Industry Conference*, 2018(1), 482–517. <https://doi.org/10.1111/1559-8918.2018.01219>
- Scheyvens, R., Banks, G., & Hughes, E. (2016). The private sector and the SDGs: The need to move beyond 'Business as Usual.' *Sustainable Development*, 24(6), 371–382. <https://doi.org/10.1002/sd.1623>
- Sells, S. P., Edward Smith, T., & Sprenkle, D. H. (1995). Integrating qualitative and quantitative research methods: A research model. *Family Process*, 34(2), 199–218. <https://doi.org/10.1111/j.1545-5300.1995.00199.x>
- Siles, I., Segura-Castillo, A., Solís, R., & Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data and Society*, 7(1). <https://doi.org/10.1177/2053951720923377>
- Sinkovics, N., Sinkovics, R. R., & Archie-Acheampong, J. (2021). The business responsibility matrix: A diagnostic tool to aid the design of better interventions for achieving the SDGs. *Multinational Business Review*, 29(1), 1–20. <https://doi.org/10.1108/MBR-07-2020-0154>
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15(1), 4–12. <https://scholarworks.uni.edu/facpub/4751>
- Starkey, K., & Madan, P. (2001). Bridging the relevance gap: Aligning stakeholders in the future of management research. *British Journal of Management*, 12(1 spec. issue.). <https://doi.org/10.1111/1467-8551.12.s1.2>
- Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press.
- Watts, D. J. (2011). *Everything is obvious: Once you know the answer*. Crown Business.
- Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, 48(2), 311–325. <http://dx.doi.org/10.1111/ejed.12014>

Authorship

André Luis Araujo da Fonseca

Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração

Rua Pascoal Lemme, n. 355, Cidade Universitária, CEP 21941-616, Rio de Janeiro, RJ, Brazil

E-mail: andre.fonseca@coppead.ufrj.br

 <https://orcid.org/0000-0002-1318-7156>

Paula Castro Pires de Souza Chimenti*

Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração

Rua Pascoal Lemme, n. 355, Cidade Universitária, CEP 21941-616, Rio de Janeiro, RJ, Brazil

E-mail: paula.chimenti@coppead.ufrj.br

 <https://orcid.org/0000-0002-6492-4072>

Maribel Carvalho Suarez

Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração

Rua Pascoal Lemme, n. 355, Cidade Universitária, CEP 21941-616, Rio de Janeiro, RJ, Brazil

E-mail: maribels@coppead.ufrj.br

 <https://orcid.org/0000-0001-9736-5273>

* Corresponding Author

Copyrights

RAC owns the copyright to this content.

Funding

The authors state that there was no financial support for the research in this article.

Conflict of Interests

The authors have stated that there is no conflict of interest.

Authors' Contributions

1st author: conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (equal); software (lead); validation (equal); writing – original draft (lead); writing – review & editing (equal).

2nd author: conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition (lead); methodology (supporting); project administration (equal); resources (lead); supervision (lead); validation (equal); writing – original draft (supporting); writing – review & editing (equal).

3rd author: conceptualization (equal); supervision (supporting); validation (equal); writing – review & editing (equal).

Plagiarism Check

RAC maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: iThenticate.

Peer Review Method

This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page, as well as the Peer Review Report, is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers and authors.

Data Availability

RAC encourages data sharing but, in compliance with ethical principles, it does not demand the disclosure of any means of identifying research subjects, preserving the privacy of research subjects. The practice of open data is to enable the reproducibility of results, and to ensure the unrestricted transparency of the results of the published research, without requiring the identity of research subjects.