DOSSIER

NEWS AGENDA GUIDED BY ALGORITMS:

content and format in *Estadão Infográficos*

DANIELA BORGES DE OLIVEIRA
Universidade Metodista de Piracicaba, Piracicaba – Brazil
ORCID: 0000-0003-2971-3840

BELARMINO CESAR GUIMARÃES DA COSTA
Universidade Metodista de Piracicaba, Piracicaba – Brazil
ORCID: 0000-0003-0808-8708

DOI: 10.25200/BJR.v16n3.2020.1284

Received on: March 26th, 2020
Desk Reviewed: May 4th, 2020
Desk Review Editors: David Domingo and Laura Storch
Revised on: October 8th, 2020
Approved on: November 3rd, 2020

ABSTRACT – This paper looks at how *Estadão Infográficos* use algorithms to examine news scripts and format content; its purpose being to use visuality both in the investigation process and for how the news is presented. We intend to demonstrate how web journalism appropriates the digital tools for innovation and differentiation. Thusly, the methodological procedure was conducted in three stages: interviewing the team members responsible for the menu, selecting those reports that use artificial intelligence techniques, and lastly, conducting a qualitative analysis upon specific concepts to explain cognitive effects that its news narratives create, which are: data visualization, image-synthesis and optical unconscious. Our investigation concluded that the use of algorithms to produce visualizations provides different perspectives of news scripts and builds communication with the public on the internet, despite the fact that this newspaper (O *Estado de S. Paulo*) does not make much use of this resource.
Key words: Data visualization. Artificial intelligence. *Estadão Infográficos*. Algorithm.

PAUTA GUIADA POR ALGORITMO: conteúdo e forma no *Estadão Infográficos*

RESUMO – O artigo observa como o *Estadão Infográficos* emprega algoritmo para prospecção de pautas e formatação do conteúdo, com o propósito do uso da visualidade tanto no processo de investigação quanto na apresentação das notícias. O objetivo é

 DOI: 10.25200/BJR.v16n3.2020.1284

demonstrar como o webjornalismo está se apropriando das ferramentas do ambiente digital para inovação e diferenciação. Assim, o percurso metodológico consiste em três etapas: entrevista a componentes da equipe responsável pelo menu, seleção de reportagens da seção que utilizam técnicas de inteligência artificial e, por fim, análise qualitativa de investigação de conceitos que explicam os efeitos cognitivos provocados pela técnica nas narrativas jornalísticas, sendo: visualização de dados, imagem-síntese e inconsciente ótico. A investigação conclui que o uso de algoritmo na produção de visualizações permite percepção de novas pautas e potencializa a comunicação com o público na internet, mesmo que tal recurso ainda seja pouco explorado pelo jornal *O Estado de S. Paulo*.
Palavras-chave: Visualização de dados. Inteligência artificial. *Estadão Infográficos*. Algoritmo.

PAUTA GUIADA POR ALGORITMO:
contenido y forma em Estadão Infográficos

RESUMEN – El artículo observa cómo *Estadão Infográficos* emplea algoritmo para prospectar pautas y formatear contenido, con propósito de usar visualidad en el proceso de investigación y en la presentación de noticias. El objetivo es demostrar cómo el periodismo web se apropia de las herramientas del entorno digital para innovación y diferenciación. Así, la trayectoria metodológica consta de tres etapas: entrevista con miembros del equipo responsable del menú, selección de informes de la sección que utilizan técnicas de inteligencia artificial y, finalmente, análisis cualitativo de la investigación de conceptos que explican efectos cognitivos provocados por la técnica en las narrativas de periodismo, siendo: visualización de datos, síntesis de imagen e inconsciente óptico. La investigación concluye que el uso del algoritmo em la producción de visualizaciones permite percepción de nuevas pautas y mejora la comunicación con el público en internet, aunque este recurso todavía es poco explorado por *O Estado de S. Paulo*.
Palabras clave: Visualización de datos. Inteligencia artificial. *Estadão Infográficos*. Algoritmo.

1 Introduction

Since its origin (1960s) the internet has been a space of interaction and simulation that allows humanity to retrieve and produce virtual messages in a decentralized way. In a flexible manner, it uses binary code (0 and 1) to translate into text, image and sound, and recombines these to generate new meanings at a rate never before seen in society. Another one of its qualities is the hypertext navigation language in complex and multisensory connections, identifiable by its non-linear constraints. Santaella (2004) refers to the immersive multilinear reader who compares, summarizes and establishes connections in constant transformation. A similar movement to that of fractal objects can be observed in cyberspace.

Despite the virtual environment having ample memory and an intense flow of information circulating within it, its effects on society and users do not necessarily provide a real understanding of its contents, which may not be perceived, or may be quickly forgotten, and thus compromising the notion of experience (Zuin & Zuin, 2015; Costa, 2002). In his cyber-flâneur allegory, Lemos (2001) describes navigating on the network as "drifting through discontinuous, fragmented and ephemeral spaces" (Lemos, 2001, pp.50–51). The observer is physically assigned to the real world while simultaneously faced with an abundance of data and virtual information. When working with social facts and structuring information, one of the biggest challenges in journalism is to ensure that messages reach the audience effectively and that those messages will make them think. This must be done before content without any verification and credibility is consumed and distributed.

Incommunication on the internet (or the failure to communicate through messages) is an issue that communication professionals should be taking note of (Wolton, 2011). Another issue they should be aware of is the disinformation process and its lack of contextualizing the facts or its production of content disguised as truth. One example is the pseudoscience propagated on social networks and websites that assume the character of unquestionable truth, and thus influence political agendas and the behavior of users. In this context it is important to observe how journalism can contribute to interpret the continuous flow of information circulating in the network and mediate educationally so that the public is capable of identifying distorted facts and false information.

In terms of adaptations in the profession, we highlight the use of specialized programs in data mining and data translation for journalistic narrative strategies such as visualizations and infographics, used by traditional media outlets that look for visual graphic solutions and content production for formats that depend on computer graphics. In this study, we analyze narratives that use artificial intelligence and data visualization from the *Estadão Infográficos*[1] menu on the digital platform of the *O Estado de S. Paulo* newspaper, also referred to as *Estadão*[2]. This is an example of a sector that appears in online portals of renowned newspapers and of strategies that alter how one thinks and conceives of news products while looking for disparity on the internet.

This requires a retrospective look in order to understand which tools the *Estadão Infográficos* uses to produce stories (mainly

  DOI: 10.25200/BJR.v16n3.2020.1284

algorithms) and the reason why this editorial is referred to as an infographic, even though it uses more formats. We chose *O Estado de S. Paulo* due to its large digital subscription base and its relevance in the use of infographics in Brazilian journalism as it was the first newspaper to include this format in its circulation. In 2019, this portal won eight awards for multimedia infographics[3] and the way it has pushed forward and adapted to a digital business model in communication is particularly noteworthy.

The methodology is based on the selection stages of the analysis *corpus*: we elected three news reports from August to September 2019 from *Estadão Infogr*áficos and applied semi-structured interviews with the former Multimedia Assistant Editor, Vinicius Sueiro, and data journalist for that section, Rodrigo Menegat. The tension between object and theoretical foundation is driven by the concepts of data visualization (Cairo, 2016), optical unconscious (Benjamin, 1987) and image-synthesis (Quéau, 1993). The categories make explicit the relationship between visuality and amplification of perception, using graphics and images to represent connections between data, which are essential for clarifying the relevance of the algorithm for discovering new journalistic guidelines that are able to hold the audience's attention in the middle of the digital distraction. This article also intends to demonstrate that the internet is not only a transitory space, but it can also have depth in narratives, usually associated to newspapers.

We contacted the portal's support section[4] in order to get a better understanding of the production process of *Estadão Infogr*áficos. Former Multimedia Assistant Editor, Vinicius Sueiro, wrote us a text message personally inviting us to visit the editorial office at the *Estadão* building on September 3, 2019[5]. The meeting (which was held from 3pm to 6pm) was a semi-structured interview with questions about the purpose of the menu, the frequency of publications, the composition of the team, the tools used, and the news scripts and innovations that *Estadão* chose to use for the web. In the occasion, we also held questions to *Estadão's* journalist, Rodrigo Menegat.

Another aspect of the article is to note the existence of different applications of algorithms for building news narratives which are not entirely dependent on automation: there are software that gather data and produce reports and complete articles at superhuman speed, with molds in journalism techniques such as StatsMonkey and

Narrative Science (Kaufman, 2019). In this study, the mathematical programming language, design and journalistic techniques are presented in an interdisciplinary fashion to build content, as we shall see below. However, it is necessary to first understand the concept of algorithm and how it has been used in journalism.

2 Algorithm: what it is and how it helps journalism

In order to understand how journalism uses algorithms as a tool, we need to know what "algorithm" means. The term is often associated with the context of interdisciplinary newsrooms for online portals, although journalists are not normally trained on how to use algorithms and data – in an interview with professionals from 149 countries, including Latin America, 79% of journalists said they wanted training in data analysis (International Center for Journalists [ICFJ], 2019). For example, the *Estadão Infogr*áficos team is composed mostly of designers and journalist who possess programming skills which are essential for producing innovative formats for the web and for data processing, according to assistant editor Vinicius Sueiro and journalist Rodrigo Menegat.

The algorithm dates back to manual application in the 9th century, with mathematician al-Khwãrizmi (Kaufman, 2019), and its definition depends on the social context (Dourish, 2016). In the digital environment: "an algorithm can be defined as a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome" (Diakopoulos, 2015, p.400). Kaufman (2019) writes: "an algorithm is a set of mathematical instructions, a sequence of tasks to achieve an expected result in a limited time" (Kaufman, 2019, p.53), which implies the definition of rules and the capacity for action. When taking into account the human factor, an algorithm can guide and influence one's perspective in content creation procedures:

> Algorithm decisions can be based on heuristics and rules, or calculations over massive amounts of data. Rules may be articulated directly by programmers, or be dynamic and flexible based on machine learning of data. Sometimes a human operator maintains agency and makes the final decision in a process, but even in this case the algorithm biases the operator's attention toward a subset of information. (Diakopoulos, 2015, p. 400).

Lindén (2017, p.62) defines an algorithm as "a self-contained step-by-step set of operations to be performed, such as calculation,

  DOI: 10.25200/BJR.v16n3.2020.1284

data processing, and automated reasoning – a set of rules that precisely defines a sequence of precise instructions that can be understood by a computer". Some of the activities it performs are prioritizing, classifying, associating, filtering and data distribution (Diakopoulos, 2015; Coddington, 2015).

For larger corporations like the "Big Five" Google, Amazon, *Facebook*, Apple and Microsoft, algorithms are a strategy to appropriate user data and personalize the content presented to the public based on search profiles. This is how algorithms are able to structure societal systems, which in turn can threaten human independence in decision-making processes. This makes what is called an "algocracy": a system "in which algorithms are used to collect, collate and organize the data upon which decisions are typically made, and to assist in how that data is processed and communicated through the relevant governance system" (Danaher, 2016, p.4). In light of this view, Danaher (2016) points out concerns about how data used by the system is concealed and the opacity of this process, something which is beyond many human beings' understanding. It is in this sense that it goes against the essential definition of transparency in data journalism and computer journalism (Coddington, 2015).

Our research addresses the use of algorithms and data visualization as strategies in the journalistic process of producing narratives. It does not focus on the complete automation that excludes the human factor in the production of news, reports, notes and articles (Lindén, 2017; Kaufman, 2019). To do this the algorithm selects information based on a set of criteria, which is another problem in itself because, despite being guided by deontology and criteria that systematize production, journalism does not establish fixed and universal rules, and organizational culture, rules and codes are constantly changing and appropriated during the production process; in other words, they are too inconsistent to write rules for an algorithm (Lindén, 2017). It would be difficult to delegate control of the entire journalistic production to the machine as it depends on human experience and analysis.

For Lindén (2017, p.64), "a journalism algorithm needs to reflect a behaviour that is not always rule-based but reflective and flexible in the sense that the application of explicit and internalized rules to complex situation is ambiguous". Having said that, guided by a journalist an algorithm can enhance human work by assisting in prospecting scripts, confirming hypotheses and observing inconsistent data.

> Traditionally, journalists have mainly relied on qualitative methods for their research, but the supply of easy-to-use, cheap, or free software for processing large data sets has opened new opportunities. The personal memory capacity and analytical processing capabilities have already been augmented with the aid of computers. (Lindén, 2017, p.64).

It is in this context that data visualization platforms use the resource to facilitate the investigation of complex databases. If you think of it as a facilitator for mining data and information units on the network, an algorithm – often linked to the use of graphics and visualizations – helps to perceive new trends, escaping the pressure of hard news.

It is no longer enough to transfer news content to the internet; this digital medium must be thought of in terms of its tools, formats and possibilities. Its routes of navigation and public interaction are much broader. Since algorithmic language is the essence of digital, the use of databases becomes the focus of the new generation of web journalism (Dalmonte, 2009), which must think about the aesthetic possibilities of online platforms, as noted by Martinez (2007, p.24): "informing well will also involve planning and design, the well thought out formatting of the 'frame' that the browser presentes. Visual and multimedia resources thus become key to new perceptions about the facts and yet still help to improve communication with the public, as we live in a visually stimulated society in which images are a universal and immediate language.

In order to use a machine to help collect, interpret, analyze and translate data a journalist needs to know certain programming techniques that will help obtain interrelationships between quantitative variables, extracted from numerical values for the perception of trends. There is vast amount of information available on the internet. According to the 2018 report from the Digital Universe Study (IDC), it is expected that by 2025 the volume of information on the internet will reach 175 zettabytes of digital information, an amount that the human cognitive capacity and its physical limitations would never be able to explore. Not only for data in digital bases, but observable characteristics such as facial expressions are recognized, quantified and organized more quickly through the use of the machine, through algorithms.

One example is the report "*O que revela uma análise das emoções dos candidatos durante o debate*"[6] (09/13/2018) – in English "What reveals an analysis of the candidates' emotions during the debate" – from *Estadão* which analyzes the facial expressions of the

  DOI: 10.25200/BJR.v16n3.2020.1284

Brazilian presidential candidates. Using four cameras all directed at their faces, Microsoft*'s* Facial Recognition Application (based on artificial intelligence that recognizes attributes in images such as facial emotions), and relying on research in the field of psychology and computer science, the team recorded percentages for the following feelings: joy, contempt, fear, anger, surprise, sadness and indifference. To do this every second of the six interviewees being filmed was transformed into an image which was subsequently uploaded to the application. This report was a collaborative effort produced by journalists and designers.

We shall explain the strategies used by journalism (data visualization, the use of algorithms, and artificial intelligence) below using multimedia reports from *Estadão Infogr*áficos as examples. For now, we need to observe the trend of this particular media outlet which, like other traditional outlets such as *Folha de S.Paulo*[7], uses infographics to create its productions, which are often not considered infographics. To do this we need to differentiate infographics from the concept of data visualization, from data-driven journalism.

3 Format editing: *Estadão Infográficos*

Working with multimedia in journalism requires knowledge of software to create graphics, interactive images, illustrations and programming language for dynamic web pages and for inserting data visualizations. Since journalists themselves do not often have the necessary knowledge required to use these tools, interdisciplinary teams are needed for complex productions in data journalism. There are a number of media outlets and menu experiences on the internet.

The latest report from the International Center for Journalists (2019) reveals a growing number of journalists interested in learning data analysis and the importance of data journalism, according to the following aspects: "data journalism helps engage audiences, improve quality and raise productivity" (ICFJ, 2019, p.10). In Brazil, the digital newspapers *Nexo*[8] and the media organization *Gênero e Número*[9] – both digital natives – are illustrative while traditional media outlets have separate sections for multimedia productions, two examples being *Infográficos* from *Folha*[10] and *Estadão Infogr*áficos. The latter, a fixed menu on the *Estadão* portal, publishes with more regularity and has its own editor.

The first Brazilian infographic "*A navegação brasileira*" – in English: *"Brazilian navigation"* – was published in the *O Estado de S. Paulo* newspaper on August 18, 1909 (Teixeira, 2010). This shows how long this outlet has been using infographics, which got its start in printed format, copied that format to use on the internet and has now been turned into digital format. Perhaps this is why the term "Infographics" is applied to most media multimedia menus that use illustrations, data, graphics and interactive resources. Formed by the inseparable combination of text and image, infographics provide an explanatory and chronological contextualization of a specific event. They are part of the informative genre since they possess the narrative logic of a text: they have a title, a lead, and name sources and authors (Teixeira, 2010). So it is said that, similar to a journalistic text, an infographic shows only a selected bit of information from a fact in order to demonstrate a specific point of view (Cairo, 2016).

In an interview with former Assistant Multimedia Editor of *Estadão*, Vinicius Sueiro[11], he outlined the structure, objectives and practices of the editorial department responsible for most of the multimedia productions for *Estadão Infogr*áficos which involve data visualization, algorithms and artificial intelligence. Editoria Multimídia, how it is called the multimedia team, is divided into sections called Photography, Layout and Infographics. Infographics section, in its turn, is divided into further sections called Template, Pure and Digital. The "Digital" section is where the team produces new formats for presenting content and data, the models of which can be reproduced in the Templates. Up until September 2019, the Digital team responsible for creating innovative formats was composed of six people: Assistant Multimedia Editor, Vinicius Souza Sueiro; visual and data journalist, Rodrigo Menegat; three web designers; and one visual journalism intern.

As Sueiro points out in his interview, *Estadão Infogr*áfico*s* is an "editorial format" rather than "editorial content", so its objective is to explore new ways of producing and presenting the news and reports. There are two ways that this can be achieved: the sector is not under the same production pressure that hard news is subjected to, meaning its agendas often differ from daily agendas; and it takes more time for the materials to be finalized, production taking anywhere from a few days to a few months.

As a page that can handle heavier products and images (and is therefore more flexible) *Estadão Infogr*áficos also takes on multimedia materials from other editorials in the newsroom which

 DOI: 10.25200/BJR.v16n3.2020.1284

use resources such as photo galleries and parallax scrolling with larger images and a background. There is no special emphasis on using data journalism as a strategy to attract audiences, since they usually access the articles from the *Infogr*áficos menu through links shared on social networks, and not directly through the website.

*Estadão Infogr*áficos news reports use tools such as Adobe Illustrator, Data-Driven Documents (D3), Datawrapper, Data Viz Project, Flourish, Python, Rawgraphics and the UVA program created by *Estadão* itself to facilitate the production of visual content and incorporate them onto the portal. We comprehend they use software to create images, graphics and aesthetic illustrations, and to create data visualizations using programming languages like Python for processing data and creating dynamic pages on the internet, or even cataloging functions and graphic types (Data Viz Project). The Data Viz Project helps suggest formats and functions of graphics to use in articles

Despite the editorial structure and the media outlet's long history of producing images, there is a lack of reference manuals or guidelines for creating infographics and data visualizations on the *Estadão* portal, and this reinforces the experimental aspect of *Estadão Infogr*áficos and its aesthetic exploration of narrative forms. Nevertheless, for the data journalist, Rodrigo Menegat, the objective of the menu is similar to that of The Pudding project. This digital outlet publishes interesting stories that are guided by mostly visual narratives: "visuals make your argument more accessible and less complex than a thousand-word essay"[12].

4 Journalism, data visualization and cognition

In order to understand the methods that the multimedia team uses for the more complex productions of *Estadão Infogr*áficos (and the fact that the outlet did not provide any web journalism references) we need to investigate the digital models of journalism so as to frame the experiences of *O Estado de S. Paulo*. According to Coddington (2015), journalistic practices are becoming more and more like programming, marked by the open code culture and the increased access to data and information sharing. In order to understand them, the author (Coddington, 2015) observes three dimensions of journalism in the digital media era: Computer Assisted Reporting (CAR), data journalism, and computer journalism.

Similar to social science techniques, statistical methods and investigative reporting, CAR provides the basis for data journalism and computer journalism, but differs from both these in terms of how it uses data in sampling, how it considers the public as a passive recipient, and being guided by the deontological principles of journalism (Coddington, 2015). Since these are more basic computing techniques, such as using digital memory or exchanging emails, the scope of CAR will not be as in-depth as the others, which are closer to the experiences found in the *Estadão*'s menu.

In use since the 2000s, data journalism can be characterized by the use of data analysis and its visual presentation, using techniques of journalistic treatment and design. By examining data we are able to discover stories and tell them in a personalized way. This is what Howard explains (2014, p.5) when he says: "in a broad sense, data journalism is telling stories with numbers, or finding stories in them"; and what Mancini and Vasconcellos (2016) say when differentiating reports produced "with" data and reports produced "from" data. Different from the former, which uses data in an illustrative and complementary way, a data report is substantiated due to the relationship obtained between the data; "the data would be the very reason for the report" (Mancini & Vasconcellos, 2016, p.75).

The following are qualities of data journalism: transparency in the production method inherited from open source culture; the exploration of large databases; and the possibility for the public to actively consider the data presented. "In developing data journalism products – often data visualization or Web – their usefulness to the audience is a prime consideration" (Coddington, 2015, p.343).

It is important to look at some of the skills that support data journalism work in a newsroom. According to Gordon (2013) there are few professionals qualified to deal with the three dimensions that are required of a data journalism team, those being: the computer-assisted reporter, who interviews the data with basic knowledge of some of the software; the developer of news applications, who knows the programming and can produce interfaces that best serve the user; and the data visualization specialist, who knows how to present data in an accessible and interesting way, and also has programming knowledge.

Cairo (2016) conceives data visualization, used in data journalism (not to be confused with CAR), as a display that allows the

  DOI: 10.25200/BJR.v16n3.2020.1284

public to explore and discover for itself as all data are present in the same representation and can be explored through navigating, thus reinforcing the active participation of the receiver:

> A data visualization is a display of data designed to enable analysis, exploration, and discovery. Data visualization aren't intended mainly to convey messages that are predefined by their designers. Instead, they are often conceived as tools that let people extract their own conclusions from the data. (Cairo, 2016).

A similar understanding is noted by journalist Rodrigo Menegat, for whom using graphics to visualize data is one way to demonstrate a quantitative, abstract event, and to make it easier for the mental operation of the reader[13]. This is what Cairo (2016) calls cognitive auxiliaries: "Graphics that encode information function as cognitive aids". Using visual representation, an individual's mental visualization of an event is representative of the image he or she is presented, thus facilitating its absorption. The idea is to prioritize in advance the aspects the brain adheres to (Cairo, 2013).

In computer journalism, public interaction is another prominent aspect: "the goal of computational journalism is to provide them [the public] with tools they need to perform their own filtering and abstraction with it" (Coddington, 2015, p.343). In this case, it is from the collection, analysis, and structuring of information that journalism performs abstraction, supported by vast databases through computation. This practice would only be possible through collaborative activities and the use of algorithms. For Lindén (2017, p.71), computational thinking is: "the integration of logical, algorithmic, scientific, and innovative dimensions of human cognition, pointing towards openness to new ideas, revolutionizing all kinds of occupations, including journalism".

This shows the proximity between data journalism and computer journalism techniques and the work at *Estadão Infogr*áficos, both in terms of structuring data with programs and algorithms and in the composition of the team. The aspect of data visualization is highlighted since the editorial department itself, formed by professionals in the fields of computing, design, and journalism, focuses on the production of visual narratives in the menu.

5 The importance of visuals

Coddington (2015) notes that abstraction in computer journalism is the cognitive ability to break up information from its material condition, such as data, that result from perceptions. By making use of visual resources, a journalist is able to abstract reflections from organizations and database comparisons. According to Cairo (2013), visualization is a type of technology that is an extension of human capacities and helps to solve problems. It is a tool called functional art because, in addition to the beauty of the imagery, it provides objectivity and accuracy through the form of structured information.

Using the distance provided by the image, the technique can be used to find previously obscure patterns by separating the information that makes up the visualization: "the purpose of all abstraction is to take distance from the concrete in order to be able to grasp it better" (Flusser, 2008, p.16). Linear text is no longer needed as the image itself can be a carrier of information, particularly the technical image (Flusser, 2008).

For data visualization, the classic format presented in reports is graphics, which gather records and demonstrate relationships between them. For current editorials which are designed to produce data-based visualizations, catalogues like *Data Viz* are widely used to understand effects and choose shapes. The gathering of data in a single visualization surface acts – as Flusser (2008) reveals about the technical image – in "attempts to bring together the punctual elements around us and in our consciousness in order to form surfaces and thereby cover the breaks" (Flusser, 2008, p.24).

In terms of infographics (term included in the *Estadão* menu) the image can be represented under the concept of image–synthesis produced by simulation in the digital environment using mathematical language and representing different abstract situations (Quéau, 1993). Journalism usually uses infographics to illustrate an event, to show how an event occurs, or to understand the structure of a particular object, particularly when photos or textual narratives cannot sufficiently describe a qualitative event.

In order to better understand the expansion of perception through visualization we can update Benjamin's concept of "optical unconscious", a term he coined at the time the film camera was invented, yet this term can still be used to explain the effects of

  DOI: 10.25200/BJR.v16n3.2020.1284

creating virtual images, such as data visualization. For Benjamin (1987), the photograph is a fixed view of a real moment that provides visual acuity to human beings, who are unable to perceive all the exact movements of nature, like the exact moment when all the legs of a horse are in the air while it's galloping.

> We are familiar with the movement of picking up a cigarette lighter or a spoon, but know almost nothing of what really goes on between hand and metal, and still less how this varies with different moods. This is where the camera comes into play, with all its resources for swooping and rising, disrupting and isolating, stretching and compressing, enlarging and reducing. It is through the camera that we first discover the optical unconscious, just as we discover the instinctual unconscious through psychoanalysis. [...] the many aspects that the recording apparatus can win from reality lie mostly outside the normal spectrum of sense perceptions. (Benjamin, 1987, pp.189-190).

In reference to media productions resulting from the logic of the cultural industry, Costa (2002) believes that messages meet pre-conformed, average standards which decidedly interfere in the aesthetic production of news. This characteristic involves the daily news agenda, the production of transitory content, and the speed of production. What is interesting about *Estadão Infogr*áficos is that it breaks from this logic due to the possibility of investigative journalism and concern with the format and visuality of the content, of dedicating itself to production with longer deadlines. Images are exactly what journalism looks to for differentiation and ways to better communicate with users.

6 Analysis: investigating the use of algorithms and visualization

Highlighting the paths that journalism has traveled in its journey to adapt to the internet and using digital technologies tools to change how investigation, information gathering, analysis, checking, and structuring of news is performed, this paper presents examples of content taken from *Estadão Infogr*áficos that use artificial intelligence, programming techniques, and design. This allows us to register techniques and solutions that can be appropriated by other media and used to train communication professionals. We look at three reports from Editoria Multimídia about data visualization, image-synthesis, and optical unconscious. These guide our analysis.

1) Data visualization and adoption in Brazil

In September of 2019, *Estadão Infogr*áficos published the article "*Simulação mostra quais crianças são adotadas (e quais não são)* no Brasil" – in English "Simulation shows which children are adopted (and which are not) in Brazil". This article received the Kantar Information is Beautiful Award in the same year in the "Best Non-English" category. The content was designed through the aid of an algorithm and data visualization, which helped view trends in the adoption of children in Brazil. However, before addressing how it was used it is important to understand how it came about.

As explained earlier, *Estadão Infogr*áficos is an editorial that creates stories with different aesthetics, narratives, and interactions with the user. To make this possible, the newsroom gives Editoria Multimídia a longer production period to work with more creative and relaxed proposals.

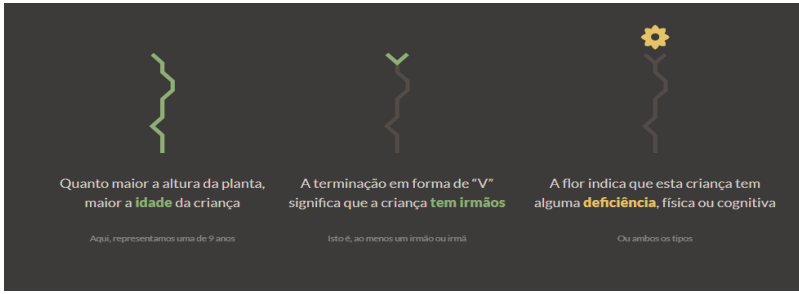
> Basically, we sorted verbs, nouns and adjectives and, in groups, we had 10 minutes to write down a few ideas that stitched those words together. We then voted on our favorites. After that, we had one week to create a raw prototype with the best ideas. The ones that seemed promising would be taken forward. The first random words my group produced had to do with fertility. We remembered a piece by the WSJ that analyzed the preferences of families in Brazil regarding sperm donors and we thought it would be interesting to compare it to the preferences regarding adoptions. As it later turned out, we had enough data to build a compelling story on the adoption process in Brazil. (Sueiro, 2019, n/p).

Using a random selection process the team combined the "fertility" category (observing the available data on the theme and the different possible approaches) with the concept of "adoption", terms that initially are not interconnected. This is why the editorial's varied features stand out and its experimentation and unpredictable combinations.

Plants were used to illustrate the fertilization category. The children available for adoption were identified according to qualities like age, race, whether they had siblings, and physical or cognitive disabilities (Figure 1). This facilitated the data visualization and made interaction more appealing due to its attractiveness.

  DOI: 10.25200/BJR.v16n3.2020.1284

Figure 1 – Legend for the child adoption simulator in Brazil.



Source: *Estadão Infográficos* (2019) [14].

The simulator is reminiscent of the National Adoption System (SNA) proposal that would be implemented in some Brazilian states as of October 2019. Using data from the Brazilian Institute of Geography and Statistics (IBGE) and the National Adoption Register (CNA), the simulator places the plants that represent children available for adoption in increasing order of age (up to 18, the age they leave shelters) with buttons for users to follow the steps of the process: a judge observes the preferences of four possible parents every day in the simulation. For adoptions, the rectangular field in the original report glows green and if the child reaches adulthood, it glows red.

Figure 2 – Adoption simulator from *Estadão Infográficos*.

Source: Estadão Infográficos (2019) [15].

In an interview with Kantar Information is Beautiful Awards, Vinicius Sueiro spoke about the tools used to produce the simulator. "The simulation algorithm was the craziest piece of software I have ever written" (Sueiro, 2019), describing how the algorithm used for the interactive was written with JavaScript. To improve the algorithm's performance, he also used the Document Object Model (DOM) – an interface that manipulates how a browser reads HTML and XML documents and identifies elements on an internet page. Sueiro (2019) says this allowed him to run the simulation thousands of times in order to reach the standards presented in the report. He adds: "The output was a JSON file that was analyzed in Python, using Pandas" (Sueiro, 2019). The Java Script Object Notation (JSON) format and the Python programming language were used through the Pandas library, and Python software being used for data analysis.

The programming knowledge that goes into the interactive simulation in all multimedia reporting is highly important. These are specific tools that require a professional who is trained in the area as most journalists themselves have not had this training. However, since the team has access to these types of tools and the expertise to use them effectively, there is greater freedom in creating the form and content to achieve innovative materials in the journalistic market.

> I believe that visual design [...] is an important factor when transforming numbers into ideas that people really feel connected to. The way in which Giorgia Lupi has covered this topic in her manifesto about data humanism is amazing – and I would like to quote a few words that really resonate with our decisions behind this story: "The more ubiquitous data becomes, the more we need to experiment with how to make it unique, contextual, intimate". (Sueiro, 2019, n/p).

In addition to the importance given to the visual aspect of the narrative, the outlet provides the codes used to produce the reports. In this case, it is an open source code which is available on the GitHub platform, a platform which has tools and projects that teach and share knowledge about software. This leads to transparency in all the team's productions.

2) Optical unconscious through Artificial Intelligence

To celebrate 200 years of the Prado Museum (Spain), the culture editor, together with *Estadão's* multimedia editor, proposed a data visualization containing all 6,367 of the museum's paintings in a report titled "*Uma inteligência artificial vai te guiar pelo Museu do Prado; nossa repórter, pela cena cultural de Madri*" (08/19/2019)

 DOI: 10.25200/BJR.v16n3.2020.1284

– in English: "Artificial intelligence will give you a tour of the Prado Museum; our reporter, for the cultural scene of Madrid". After extracting mathematical descriptions of each work (the features) the team compared these values and calculated the distance between the paintings marked by their similarities. They used the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm to do this, which allowed each work to be featured on a single surface/image.

Figure 3 – Visualization of more than 6 thousand works at the Prado Museum.

Source: *Estadão Infográficos* (2019) [16].

In order to reinforce the transparency of the method to the public, the report provides the source code used for the story as well as the image representing the works in a single view, in this way the reader can enlarge and explore it. Grouping the paintings would be a highly improbable task, maybe even impossible, without the aid of a computer as recognition using just the human eye would take so much longer; the eye is subject to physical obstacles such as the distance from the rooms in which the works are exhibited. Visualization, however, broadens the perception, making it easier to see which painters created works that shared similar characteristics to each other, even ones spread out over time. It is able to observe each artist's styles, their characteristics, aesthetic proposals and criticisms of their time.

The potentiation of the human eye corresponds to the effect that Benjamin (1987) describes as the optical unconscious, that

photographic images broaden the field of vision to include features that the eye cannot directly observe. Using illustrations constructed by the mathematical values that express the aesthetic similarities of each painting, abstracting interrelationships between the works, understanding trends and even the limits of AI become clearer. In this case, the machine understood that very similar works, such as "*La maja vestida*" by Francisco de Goya (1800) and "*La maja desnuda*" by Francisco de Goya (1795) were one and the same: as a result, the visualization of these works is superimposed on each other. This excerpt from the story explains this: "The computer is unable to understand the historical context or the subjective characteristics of a work [...]. In other words: it is up to us, humans, to give the script greater meanings and relationships" (Toni et al., 2019).

Despite some limitations with the machine, the cloud-shaped image shows the inclination in the work of some artists: Carlos de Haes' works (1829–1898) were grouped together due to the similarities between them, while Francisco de Goya's paintings (1746–1828) were more spread out, showing a variety in his style. It is interesting to note that the choice to use illustrations as a form for storytelling in the report, in addition to allowing new ways of evaluating the museum's paintings, proposes an "interesting way to see the institution's collection" (Toni et al., 2019), an excerpt that explains the search for innovation and curiosity in the construction of content.

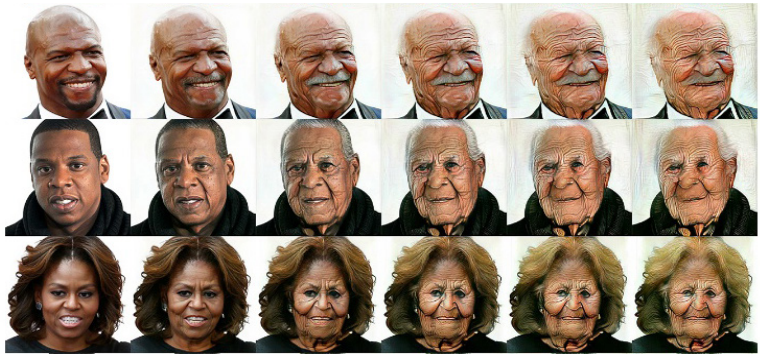
3) Image-synthesis: simulation and journalism

Simulated images are widely used in journalism, especially when there is a lack of image resources like photos of reported events to depict the event. A different case was the use of the FaceApp application in journalism, a tool that runs filters to rejuvenate or age a person's face in order to find manipulations that reinforce the use of biased data sources.

The article "*Além de envelhecer, FaceApp embranquece rostos negros*" (08/02/2019) – in English: "In addition to aging, FaceApp whitens black people's faces", demonstrated how biased databases perpetuate the current problem of standards of beauty and prejudice against black people. For this, the multimedia team tried applying the aging filter five times to the same photo. The result, in addition to whitening faces, arises from the exploration of images that have no reference in the real world, that is, they are configured as abstract languages.

DOI: 10.25200/BJR.v16n3.2020.1284

Figure 4 – Aging filter on FaceApp applied five times to black people's faces.



Source: *Estadão Infográficos* (2019) [17].

The perspective of Editoria Multimídia, which seeks to present interesting stories with innovative visual narratives, is summarized in the following excerpt: "Outside of a few curious people, no one runs the filter five times on the same image" (Romani et al., 2019) to see how the algorithm performs. The editorial department has a laboratory profile which operates in a similar way to the scientific methodology: there is a hypothesis, a method used and explained to the public, with experimentation and observable results through visualizations. This corroborates Wolton's (2011) conception of the importance of journalists and scientists as intermediaries in the effective communication with society.

The simulated image is subject to distortions as it is produced by a machine which uses a database to perform the operation. In this case, the error of presenting more images of white aged faces to the computer than black is a human one; the machine is trained to understand that color is also a criterion for aging. This is what characterizes the machine learning technique, a machine performing an operation based on organized data to teach it. This same type of technology is used in facial recognition.

The human factor stands out in the data feed for the application to run the filters, and in the reflection on the results of these images by Editoria Multimídia. It is not the case then that using the machine is sufficient to understand the phenomenon; journalists need to clarify the problem through the techniques of the profession: checking the information with sources and seeking contextualization. For example, the report informs the reader about past errors involving FaceApp with its hotness filter application which, in 2017, made

user photos appear more attractive by whitening their faces. If no explanation is given, the aging filter may appear like one punctual problem. Without the use of illustrations in the report it is difficult to fully understand the whitening effect, which is precisely the focus of the article. Here, the image produced by the machine is the theme of the content to demonstrate the use of biased databases. The procedure is known to users as the application was popular at the time and many people had it on their cell phones; thus, the sequence of images was part of the reader's repertoire, they were accustomed to the format of the application. It is about clarifying with images, but also about bringing the public closer to the reporting procedure.

7 Final considerations

Our results show that *Estadão Infogr*áficos reinforces freedom from the daily agenda by choosing guidelines and producing content, the result of unexpected connections of concepts, data and information from the use of software and algorithms. This is enhanced by the ubiquity of information available on the network. In this context, curiosity is a strategy to propose narratives not only in determining or defining an agenda, but in the way the stories are presented and how to make them more interesting to the public by adding new aesthetic possibilities.

Experimentation equipped with data visualization and algorithm operations that act as cognitive aids for journalists to abstract perceptions is essential. New guidelines emerge from computing combined with journalistic techniques and possible forms of presentation – which *Estadão Infogr*áficos does in its search for innovative formats to compose models. Computer thinking offers new ways of looking at social events and provides a different approach to news and reports: journalists trained in traditional techniques can adapt to the daily production of news which requires speed, fragmentation and accuracy, but impairs the checking and contextualization of facts. The autonomy of publishing infographics and data visualization and the diversification of techniques use creativity to circumvent barriers of traditional newsrooms by customizing complex context that enhance the public's perception of exploration and discovery.

The interdisciplinary teams, made up of journalists, designers and programmers, have reinvigorated the current model of multimedia journalism. This demonstrates a trend in applying data journalism to productions, which is a result of adapting to the characteristics of

  DOI: 10.25200/BJR.v16n3.2020.1284

digital media, associated with language hybridization and a multiplicity of connection points and content access in mobile and interactive structures. There is a predisposition for this observed in digital natives and in sections of traditional newspapers such as *Estadão Infogr*áficos.

Lastly, the research shows that concepts from the social sciences, such as image–synthesis and the optical unconscious, can be updated to support investigations of visual journalism as an element of functional art to be produced on the internet. In this case we studied, the creation of data visualizations and simulated images by algorithms favor cognitive processes of content and memory assimilation, also affecting the formation of new narratives, perceptions and experiences.

NOTES

1 "Infographics" tab located in options on the left side of the menu on the *Estadão* website, or *O Estado de S. Paulo*. Retrieved from www.estadao.com.br/infograficos

2 Printed newspaper founded in 1875 under the title *A Província de São Paulo*. After the Proclamation of the Brazilian Republic in 1890 it changed its name to *O Estado de S. Paulo*, also known today as *Estadão*. In 2000 the domain www.estadao.com.br was created. According to the Communication Verifier Institute (IVC), *Estadão* is one of the media outlets with the largest print and digital circulation in Brazil. Its site has had more than 31 million visits. Retrieved from http://patrocinados.estadao.com.br/medialab/about-me/

3 Retrieved from https://brasil.estadao.com.br/noticias/geral, estado-ganha-8-premios-em-concurso-de-design-veja-as-reportagens,70003049883

4 Contact made on August 19, 2019 via email, to the following electronic addresses: falecomestado@estadao.com and portal@estadao.com.

5 Interview with *Estadão's* Multimedia Assistant Editor, Vinicius Sueiro, and *Estadão's* data journalist, Rodrigo Menegat, granted to journalist Daniela Borges de Oliveira on September 3, 2019 at the *O Estado de S. Paulo* headquarters (Av. Eng Caetano Álves, 55, Limão/SP). The dialogue is included in the monographic project "*Narrativa Digital: Análise Crítica das Estratégias do Estadão Infográficos*" (in English, "Digital Narrative: A Critical Analysis of Estadão Infog*ráficos* Strategies"). Mentored by Professor Belarmino Cesar Guimarães

da Costa and presented to the journalism course at the Methodist University of Piracicaba (UNIMEP) to obtain a bachelor's degree.

6 Retrieved from www.estadao.com.br/infograficos/politica,o-que-revela-uma-analise-das-emocoes-dos-candidatos-durante-o-debate,923037

7 *Folha de S.Paulo, Globo* and *Estadão* are the media outlets with the largest national print and digital circulation in Brazil. The *Folha de S.Paulo* newspaper was founded in 1960 and its website is: www.folha.uol.com.br

8 *Nexo:* Native digital newspaper launched in 2015 to produce in-depth content based on data and statistics. Retrieved from www.nexojornal.com.br

9 *Gênero e Número*: news portal that provides data on gender equity. Retrieved from www.generonumero.media

10 The "Infográficos" section of *Folha* accumulates multimedia productions on its site www1.folha.uol.com.br/infograficos/

11 As discussed in the methodological procedures, the author conducted a semi-structured interview with former Multimedia Assistant Editor, Vinicius Souza Sueiro, at 3pm on September 3, 2019 at the *O Estado de S. Paulo* headquarters. Data journalist, Rodrigo Menegat, was also interviewed on this day.

12 Retrieved from https://pudding.cool/about/#pitch

13 Concept expressed in module 4 of the course "*Introdução ao jornalismo de dados: como entrevistar dados para reportagens investigativas*" [Introduction to data journalism: how to interview data to investigative reporting], available on the platform JournalismCourses.org from the Knight Center for Journalism in the Americas (August 5 to September 8, 2019). Check: https://journalismcourses.org/pt-br/course/dados/. This concept was reinforced during an interview on September 3, 2019.

14 Retrieved from https://arte.estadao.com.br/brasil/adocao/criancas/

15 Retrieved from https://arte.estadao.com.br/brasil/adocao/criancas/

16 Retrieved from www.estadao.com.br/infograficos/viagem,uma-inteligencia-artificial-vai-te-guiar-pelo-museu-do-prado-nossa-reporter-pela-cena-cultural-de-madri,1025741

DOI: 10.25200/BJR.v16n3.2020.1284

17 Retrieved from www.estadao.com.br/infograficos/link,alem-de-envelhecer-faceapp-embranquece-rostos-negros,1018384

REFERENCES

Benjamin, W. (1987). A obra de arte na era de sua reprodutibilidade técnica [The work of art in the age of its technical reproducibility]. In W. Benjamin (Ed.), *Magia e Técnica, Arte e política: Ensaio sobre literatura e história da cultura* (3ª ed.) (pp.165–196). São Paulo: Editora Brasiliense.

Cairo, A. (2013). *The Functional Art: An introduction to information graphics and visualization*. Berkeley: New Riders.

Cairo, A. (2016). *The Truthful Art: Data, charts, and maps for communication*. Berkeley: New Riders.

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: a typology for evaluation data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3 (3), 331-348. DOI: 10.1080/21670811.2014.976400

Costa, B. C. G. da. (2002). *Estética da violência: Jornalismo e produção de sentidos* [Aesthetics of violence: Journalism and production of meanings]. Campinas: Autores Associados, Piracicaba: Editora Unimep.

Dalmonte, E. F. (2009). *Pensar o discurso no webjornalismo: temporalidade, paratexto e comunidades de experiência* [Thinking discourse in webjournalism: temporality, paratext and communities of experience]. Salvador: EDUFBA.

Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, 29, 245–268. DOI: 10.1007/s13347-015-0211-1

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3 (3), 398–415. DOI: 10.1080/21670811.2014.976411

Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, 3(2), 1–11. DOI: 10.1177/2053951716665128

Flusser, V. (2008). *O universo das imagens técnicas* [The universe of technical images]. São Paulo: Annablume.

Gordon, R. (2013, Jun. 28). Want to build a data journalism team? You'll need these three people. *Knight Lab*. Retrieved from knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/

Howard, A. B. (2014). *The Art and Science of Data-Driven Journalism*. Tow Center for Digital Journalism: Columbia Journalism School. DOI: 10.7916/D8Q531V1

International Center for Journalists (ICFJ). (2019). *The State of Technology Global Newsrooms*. Retrieved from www.icfj.org/sites/default/files/2019-10/2019%20Final%20Report.pdf

Kaufman, D. (2019). A Inteligência Artificial mediando a comunicação: Impactos da automação [Artificial Intelligence mediating communication: Impacts of automation]. In M. Barbosa (Ed.), *Pós-verdade e fake news: Reflexões sobre a guerra de narrativas* (pp.49–58). Rio de Janeiro: Cobogó.

Lemos, A. (2001). Ciber-flânerie. In D. F. Silva, & S. Fragoso (Eds.), *Comunicação na Cibercultura* (pp.45–60). Leopoldo: Editora Unisinos.

Lindén, C. (2017). Algorithms for journalism: The future of news work. *The Journal of Media Innovations 4.1*, 4 (1), 60–76. DOI: 10.5617/jmi.v4i1.2420

Mancini, L., & Vasconcellos, F. (2016). Jornalismo de dados: Conceito e categorias [Data journalism: Concept and categories]. *Revista Fronteiras,* 18 (1), 69–82. DOI: 10.4013/fem.2016.181.07

Martinez, A. G. (2007) A construção da notícia em tempo real [Building news in real time]. In P. Ferrari (Ed.), *Hipertexto hipermídia: as novas ferramentas da comunicação digital* (pp.13–27). São Paulo: Contexto.

Quéau, P. (1993) O tempo do virtual [The time of the virtual]. In A. Parente. (Ed.), *Imagem-máquina: a era das tecnologias do virtual* (pp.91–99). Rio de Janeiro: Editora 34.

Romani, B., Ponceano, B., & Sueiro, V. (2019, Mar. 19) Além de envelhecer, FaceApp embranquece rostos negros [In addition to aging, FaceApp whitens black people's faces]. *Estadão.* Retrieved from www.estadao.com.br/infograficos/link,alem-de-envelhecer-faceapp-embranquece-rostos-negros,1018384

Santaella, L. (2004). *Navegar no ciberespaço: O perfil cognitivo do leitor imersivo* [Navigating cyberspace: The cognitive profile of the immersive reader]. São Paulo: Paulus.

Sueiro, V. (2019). Interview with 2019's Best Non-English-Language Winner. *Kantar Information Is Beautiful Awards*. Retrieved from www.informationisbeautifulawards.com/news/540-interview-with-2019-s-best-non-english-language-winner

Teixeira, T. (2010). *Infografia e Jornalismo*: *Conceitos, análises e perspectivas* [Infographics and Journalism: Concepts, analyzes and perspectives]. Salvador: EDUFBA.

Toni, B., Menegat, R., Marin, C., & Ponceano, B. (2019, Aug. 19). Uma inteligência artificial vai te guiar pelo Museu do Prado [Artificial

  DOI: 10.25200/BJR.v16n3.2020.1284

intelligence will give you a tour of the Prado Museum; our reporter, for the cultural scene of Madrid]. *Estadão.* Retrieved from www.estadao.com.br/infograficos/viagem,uma-inteligencia-artificial-vai-te-guiar-pelo-museu-do-prado-nossa-reporter-pela-cena-cultural-de-madri,1025741

Wolton, D. (2011). *Informar não é comunicar* [Informing is not communicating]. Porto Alegre: Sulina.

Zuin, A. A. S., & Zuin, V. G. (2015). A indústria cultural, a semiformação e a revolução microeletrônica [The cultural industry, semiformation and the microelectronic revolution]. In A. F. Maia, A. A. S. Zuin, & L. A. C. N Lastória (Eds.), *Teoria Crítica da cultura digital*: *Aspectos educacionais e psicológicos* (pp.202–213). São Paulo: Nankin.

DANIELA BORGES DE OLIVEIRA. Journalist at the Methodist University of Piracicaba (UNIMEP), she was a PIBIC/CNPq scholarship holder for the Scientific Initiation Study "*Folha de S.Paulo*: News coverage, aesthetic violence and media society" (August 2019 to July 2020), student in the research group "Comunication in the Digital Age: Culture, Aesthetics and Language" (UNIMEP). Research areas: Digital Culture, Cultural Industry and Journalism Theory. E-mail: dani7b.o@gmail.com

BELARMINO CESAR GUIMARÃES DA COSTA. Holds a PhD and Master's Degree in Education (Unicamp/UFSCar) and is also a journalist (UNIMEP). Professor of the Graduate Program in Education at UNIMEP. Develops research on media, language and education; cultural industry and digital environment; communication, technology and aesthetics. Member of the graduation program "Critical Theory and Education" (UFScar). Former coordinator of the research group "Communication and Education" at Intercom. E-mail: belarminocesar@uol.com.br

TRANSLATED BY: LEE SHARP